Exploring the Functional Role of SETD2/H3K36me3 during Cellular Differentiation

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Christina Ambrosi

aus

Deutschland

Promotionskommission

Prof. Dr. Tuncay Baubec
(Vorsitz und Leitung der Dissertation)

Prof. Dr. Matthias Altmeyer

Dr. Johannes Zuber

Zürich, 2020

Mathematisch-naturwissenschaftliche Fakultät der Universität Zürich

Dissertation

Exploring the Functional Role of SETD2/H3K36me3 during Cellular Differentiation

Autor

Christina Ambrosi

Department of Molecular Mechanisms of Disease

Universität Zürich

Winterthurerstrasse 190

8057 Zürich, Schweiz

I. Summary

Chromatin modifications on DNA and histones, are dynamic and reversible marks required for genome function and transcriptional regulation by altering genome organisation and accessibility. They are relevant for mammalian development and disturbed patterns are associated with human disease. The mechanisms that deposit and regulate chromatin modifications have been largely elucidated. However, the current challenge remains to understand how these patterns precisely influence gene activity and other biological processes.

In particular, the trimethylation of histone H3 at lysine 36 (H3K36me3) is suggested to play an important role in active gene transcription. This modification accumulates on the bodies of all highly-transcribed genes through a mechanism which is highly conserved from yeast to human. This post-translational histone modification is uniquely set by the RNA Polymerase II-interacting methyltransferase SETD2. Impaired action of this enzyme and the consequent loss of H3K36me3 histone marks are increasingly reported to be key events in promoting cancer growth and malignancy. Thus, gaining insights into their role in mediating proper chromatin and cellular function could provide essential contributions to human health. Based on yeast studies, it has been found to be involved in regulating transcriptional elongation and mRNA processing. However, the function of SETD2/H3K36me3 in mammalian systems remains poorly understood.

Towards understanding the function of SETD2-mediated H3K36me3 in mammalian cells, we generated murine embryonic stem cells devoid of H3K36me3 as a system that enabled us to functionally address the influence of the enzyme and its mark on gene regulation. Comprehensive, genome-wide quantification of transcriptome and interactome differences by functional genomics methods allowed to identify a crucial role for SETD2/H3K36me3 in establishing gene expression programs during neuronal differentiation. We could further identify a potential mechanistic link between these transcriptome changes and reduced recruitment of splicing regulators to the elongating polymerase.

II. Acknowledgement

I would like to express my appreciation towards those who have helped and supported me in completing my PhD research over the last 4 years. First and foremost, I am grateful for my supervisor Prof. Tuncay Baubec. He did not only provide me with the opportunity to conduct this exciting research project at hand, but also supported me with his expert scientific guidance and continuous encouragement.

I would like to further express my deep gratitude to my PhD committee members Dr. Johannes Zuber, Prof. Matthias Altmeyer, Prof. Csaba Földy and Prof. Lukas Pelkmans for providing me with their valuable and constructive suggestions for the advancement of this research work. The willingness to give their time generously has been very much appreciated.

I would also like to thank the staff of the UZH core facilities (Functional Genomics Center Zurich, Cytometry Facility) for their assistance and all collaborators for their support whereby this work would not have been complete.

I am especially grateful for the Baubec lab members who have also become my friends along the way. I (aka Chromatina) could always count on their support, both scientifically and personally. First, my appreciation goes to Dr. Massimiliano Manzo (aka Silver Server) for sharing the time as first PhD students in the lab, for making me laugh from the beginning and for patiently teaching me the depths of bioinformatics. I would like to thank Joël Wirz (aka The Bioruptor) for his precise hands on library preparations and Stefan Butz (aka ATAC Attack) for his scientific discussions as well as hilarious stories over coffee and for making long evenings in the lab entertaining. I would also like to thank Dr. Nina Schmolka for her cheerful nature and for giving me strength and self-awareness by holding her amazing Yoga lessons. I am also thankful to Dr. Rodrigo Villasenor for his scientific advice and for introducing me to the open science network. A special thank you goes to Ramon Pfändler for acknowledging me as a “carrot and stick” supervisor and for his outstanding help with mass spectrometry procedures, and to Dr. Ino Karemaker for keeping up the organisation in the lab. Last, but not least I would like to show my appreciation to Davide Recchia for patiently discussing experimental bottlenecks arising from working with (way too) large proteins. I wish everyone of you the best of luck for your future with the hope to stay in touch, independent of time or distance.

I would also like to express my appreciation to all the people of the DMMD for contributing to an encouraging and motivating working atmosphere. I am especially grateful to the following ladies for

their understanding, guidance, laughter and happiness: thanks to the fun and loving Jeanette Abplanalp, to Ann-Katrin Hopp, the most empathic and thoughtful person I know, to the kind Lavinia Bisceglie and to the cheerful Kathrin Nowak. A huge thank you also goes to Zyanya Diaz Hirashi for always having an open ear and for sharing her infinite positive Mexican spirit.

I would also like to thank my beloved friends Eva Hock and Christopher Gassan for welcoming me to Zurich and the most amazing people from home Josefine Gierth, Christin Kopenhagen, Katharina Rossow, Bea Herz, Sabrina Grasse, Juliane Senf, and Kai Skrodczky for being there with me for all the happy adventures, but also misfortunes along the way. I am also extremely grateful to the charming and warm-hearted Rodrigo Peña Hernández and Emeric Monnet for all our travel as well as wine adventures and all the exquisite and fun dinner nights in Elliot's kitchen.

I would like to express my deepest appreciation to a very special person in my life: Eric Schulz, my partner in crime who stood by my side through this adventure, supporting me in every way possible and showing me that distance does not matter when it comes to love. I am also immensely thankful to Jamie Ambrosi and Laura Riolo for being the best sisters by heart one can wish for.

Lastly, I am forever grateful to my parents Andrea und Horst-Dieter Ambrosi, my brother Thomas Ambrosi and to the rest of my family: Ihr seid und bleibt für mich mit eurer endlosen Liebe und Unterstützung die größten Vorbilder und ich werde euch auf ewig dafür dankbar sein, dass ihr mich zu der Person gemacht, die ich heute bin.

III. Table of Content

IV. Abbreviations

5-hmC	5-hydroxymethylcytosine
2i	two inhibitors PD03259010 and CHIR99021
53BP1	p53-binding protein
5mC	5-methylcytosine
AEBP1	AE binding protein 2
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
ATAC-seq	assay for transposase-accessible chromatin using sequencing
ATM	ataxia telangiectasia-mutated
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and rad3-related
bp	base pairs
CAF1	chromatin assembly factor 1
CAGE	cap analysis gene expression
CBP	CREB-binding protein
CBX	chromobox protein homolog
ccRC	clear cell renal cell carcinomas
CDK	cyclin-dependent kinase
CENP-A	histone H3-like centromeric protein A
ChEP	chromatin enrichment for proteomics
ChIP	chromatin immunoprecipitation
CPSF	cleavage-polyadenylation specificity factor
CRABP1	cellular retinoic acid-binding protein
CstF	cleavage stimulatory factor
CTD	carboxy-terminal domain
CtIP	carboxy-terminal binding protein and interacting protein
DBD	DNA-binding domain
DDR	DNA damage response
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxynucleoside triphosphate
DOC	deoxycholate
DOT1L	disruptor of telomeric silencing 1-like
DRB	5,6-Dichloro-1-β-d-ribofuranosylbenzimidazole
DSB	DNA double-strand break
DSIIF	DRB sensitivity inducing factor
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid

EGTA	ethylene glycol tetraacetic acid
eIF3	eukaryotic translation initiation factor 3
FACS	fluorescence-activated cell sorting
FACT	facilitates chromatin transcription
FGF4	fibroblast growth factor 4
G2 phase	gap 2 phase
GTF	general transcription factor
GTP	guanosine-5'-triphosphate
H2AK119ub1	histone H2A lysine 119 monoubiquitination
H3K36me3	trimethylation of Lysine 36 on Histone 3
HAT	histone acetyltransferases
HIRA	histone cell cycle regulation-defective homologue A
hnRNP	heterogeneous nuclear ribonucleoprotein
HP1	heterochromatin protein 1
HR	homologous recombination
KAP1	krüppel-associated box
kb	kilobases
KMT	histone lysine methyltransferase
KRAB	KRAB associated protein 1
LEDGF	lens epithelium-derived growth factor
LIF	leukemia-inhibitory factor
LIG4	DNA ligase IV
m6A	N6-methyladenosine
m7G	7-methylguanylate
MDC 1	mediator of DNA damage checkpoint protein 1
mES	mouse embryonic stem
METTL	methyltransferase like
MgCl2	magnesium chloride
MLL	myeloid/lymphoid or mixed-lineage leukemia 1
MRN	MRE11-RAD50-NBS1
mRNA	messenger RNA
MS	mass spectrometry
MSH6	MutS Homolog 6
NaCl	sodium chloride
NELF	negative elongation factor
NER	nucleotide excision repair
NEUROD1	neurogenic differentiation factor 1
NEUROG/Ngn	neurogenin
NGD	no-go decay
NHEJ	non-homologous end joining
NMD	nonsense-mediated decay

NP-40	nonidet P40
NPC	nuclear pore complex
NP	neuronal progenitor
NSD	non-stop decay
NSD1	nuclear receptor-binding SET domain protein 1
NuRD	nucleosome remodelling deacetylase
OLIG2	oligodendrocyte transcription factor 2
PABP	poly(A)-binding protein
PAF1	polymerase-associated factor 1
PAGODA	pathway and gene set overdispersion analysis
PAS	polyadenylation site
PAX	paired box protein
PCL	polycomb-like protein
PCNA	proliferating cell nuclear antigen
PCR	poly chain reaction
PHC	polyhomeotic-like protein
PHD	plant homeodomain
PIC	preinitiation complex
Pol II	RNA Polymerase II
POU5F1	POU domain, class 5, transcription factor 1
PRC	polycomb repressive complex
pre-mRNA	premature mRNA
PTB	polypyrimidine tract-binding protein
PTM	post-translational modification
RBAP	retinoblastoma-binding proteins
RBP	DNA-directed RNA polymerase II subunit
RING	ring finger protein
RNA	ribonucleic acid
RQC	ribosome quality control
rRNA	ribsosomal RNA
RT-qPCR	quantitative reverse transcription PCR
S phase	synthesis phase
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
scRNA-seq	single-cell RNA sequencing
SDS	sodium dodecyl sulfate
SEC	super elongation complex
SET	su(var)3-9, enhancer of zeste, trithorax
SETD2	SET domain containing 2
SILAC	stable isotope labelling with amino acids in cell culture

snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoproteins
SRI	Set2 Rpb1 interacting domain
ssDNA	single-stranded DNA
SWI/SNF	SWItch/sucrose non-fermentable
TBP	TATA-binding protein
TC-NER	transcription-coupled
TE	Tris-EDTA
Tet	tetracycline
TF	transcription factor
TN	terminal neuron
TREX	transcription-export complex
TRIM28	tripartite motif-containing 28
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TSS	transcriptional start site
TTS	transcriptional termination site
UHRF1	ubiquitin-like, containing PHD and ring finger domain 1
UMAP	uniform approximation and projection method
UPF1	regulator of nonsense transcripts 1
WDR33	tryptophan-aspartic acid (W-D) Repeat Domain 33
WT	wild-type
WTAP	wilms tumor 1 associated protein
XLF	XRCC4-like factor
XRCC4	X-ray repair cross-complementing protein 4
XRN2	5'-3' exoribonuclease 2
YTHDC1	YTH-domain containing 1
YTHDF	YTH domain-containing family protein
ZF	zinc finger
ZMYND11	zinc finger MYND domain-containing protein 11

1. Introduction

1.1. Gene Regulation

Each cell within a multicellular organism has distinguishable properties. This specific cellular identity is determined by the expression of genes in a time and place-dependent manner and is passed onto daughter cells by sequence-dependent and -independent mechanisms. Thus, the governance of gene expression is a highly complex and tightly-regulated process that ensures cell identity at any given time.

1.1.1. Chromatin

The central dogma of molecular biology defines the framework for the faithful information transfer between biopolymers, comprising DNA, RNA, and polypeptides, in living organisms. Their sequences encode biological information which can be copied, converted, and translated in an irreversible manner. Cellular replication is the process of copying the double-stranded DNA molecules during S phase of the cell cycle, assuring correct information transfer to daughter cells upon cell division. Transcription is the process of converting a specific sequence of DNA into RNA in the nucleus. Translation is the final process of decoding messenger RNA molecules into functional polypeptides and proteins in the cytoplasm of the cell. Further levels of regulation may occur through co- or post-transcriptional and -translational processes (Crick, 1970).

The human body consists of over 200 different cell types which exhibit particular characteristics and functions whilst containing the same genetic material in the nucleus - the DNA. The smallest units of the latter are nucleotides which consist of three different chemical moieties: a backbone of a phosphate group and a five-membered deoxyribose sugar ring, completed with a specific nitrogen base. In DNA, there are four types of nitrogen bases: the pyrimidines cytosine (C) and thymine (T) as well as the purines adenine (A) and guanine (G). Their specific order determines all further information transfer and is thereby referred to as the genetic code. Specifically, a three-nucleotide codon in a nucleic acid sequence specifies a single amino acid, the smallest unit of proteins. In its biological form, DNA forms a right-handed double-helix in which two antiparallel strands are coiled around each other via hydrogen bonds formed between a purine and pyrimidine

(base pairs A-T and G-C). This structure exhibits a minor and a major groove of which the latter is wider and more frequently subject of DNA binding proteins (Alberts, 2002).

The human genome consists of roughly three billion base pairs which have to be assembled in the approximately 6 µm-wide diameter of the nucleus. This requires a high degree of compaction realised through a complex structure of DNA and proteins - the chromatin. The fundamental element of chromatin is constituted by the nucleosome, a core particle of 146 DNA base pairs wrapped around an octamer of basic histone proteins: two dimers of histones H2A and H2B as well as a tetramer of histones H3 and H4 (Kornberg and Thomas, 1974). Proteins which assist nucleosome assembly and folding into an oligomeric structure are called chaperones (Ellis, 2006). Histones are highly conserved, small proteins with a central fold domain consisting of three alpha-helices separated by two loop regions, constituting a positively charged protein motif recognising specific DNA sequences (Arents et al., 1991; Luger et al., 1997). For instance, the negatively charged DNA double helix is bent in 1.65 tight turns around the histone octamer which requires extensive compression of the minor groove. AT-periodic sequences are easier to compress and therefore found more frequently in segments where the minor groove faces the octamer core, facilitating a sharper bending of the DNA around the nucleosome. In contrast, nucleosome linker regions show strong preference to sequences that resist DNA bending and do not aid nucleosome formation (Struhl and Segal, 2013).

The histones' N-terminal tails are variable and rather unstructured, but generally rich in lysine and arginine residues - the sites of various post-translational modifications (PTMs) (Bannister and Kouzarides, 2011). In addition to sequence composition, these PTMs can change the overall charge of the histone core which facilitates alterations in DNA accessibility and protein-protein interactions with the nucleosome, thereby playing a key role in gene regulation (Lawrence et al., 2016). The chapter 1.2.2. focuses in more detail on of post-translational modifications of histones and their regulatory role.

Several nucleosomes can form a nucleofilament by their regular spacing along the genome. This in turn can adopt series of more complex structures, such as 30 nm fibres, ultimately resulting in highly-condensed, coiled chromosomes. With the exception of the germ cells, each human cell contains a total of 46 chromosomes - two copies of each of the 22 homologous chromosomes, one set inherited from the mother and one from the father, plus two so-called sex chromosomes (X and Y in males, two Xs in females). Besides the lack of space in the nucleus, the different forms of compaction are also necessary to regulate gene expression as well as the replication of the DNA molecule during cell division. This latter is facilitated through an ordered series of stages,

collectively referred to as the cell cycle. Therein, the coiled chromosomes are highly condensed and subsequently separated for distribution to the two daughter cells during mitosis, the division of the nucleus of a eukaryotic cell.

During interphase, the most frequent, metabolic cell cycle phase (including G1, S, and G2 phase), when the cell is not dividing, the DNA molecules are highly transcribed and replicated. For these processes, parts of the chromatin need to be accessible and uncoiled which is defined as euchromatin. On the contrary, heterochromatin does not alter its condensation throughout the cell cycle and is formed principally on the periphery of the nucleus (Rae and Franke, 1972; Passarge, 1979). Thus, the dynamic structure of chromatin results in functional territories, permitting or restricting its accessibility and interaction with protein machineries, which ultimately influences all processes of gene regulation (Dixon et al., 2016).

1.1.2. Transcription Factors

In a given cell type, only a subset of genes are expressed in a highly-regulated fashion with many levels and stages of control. Transcription factors (TFs) can decode the genome by interpreting DNA sequences and exerting control over gene expression which determines cell identity (Lee and Young, 2013). Their regulatory domains and physiological roles are highly conserved among metazoans. However, over time many TFs have diverged with growing complexity and control different regulatory gene networks depending on the cell type (Bejerano et al., 2004). The number of TFs scales with genome size (van Nimwegen, 2003). Roughly 10 % of genes in the genome encode TFs, comprising the largest family of human proteins (Fulton et al., 2009; Vaquerizas et al., 2009).

TFs bind DNA with their DNA-binding domain (DBD) in a sequence-dependent manner by recognising specific motifs at cis-regulatory sites (promoters and enhancers) adjacent to the genes they regulate (Mitchell and Tjian, 1989; Ptashne and Gann, 1997). By binding to these regulatory sites, they initiate or restrict transcription by impacting the chromatin state through attraction or repulsion of other factors. Further, the binding of several, distinct TFs (homotypic clusters) is often required for a cooperative action in gene regulation (Jolma et al., 2015).

The recognition of similar motifs due to similarity in their tertiary structure typically corresponds to specific TF families or subfamilies, like C2H2 zinc finger (ZF) or Helix-loop-helix factors (bHLH), which are differentially expressed among different cell types (Matys et al., 2006). On the contrary,

general transcription factors (GTFs, e.g. TFIIA, TFIID) are ubiquitously expressed, bind to the region surrounding the transcriptional start site of a gene, referred to as core promoter, and facilitate the formation of a preinitiation complex (PIC) upon transcriptional activation.

1.1.3. Eukaryotic Transcription

Eukaryotic transcription is the process of converting a specific sequence of DNA into RNA in the nucleus and proceeds in three steps: initiation, elongation, and termination (Figure 1).

1.1.3.1. Transcriptional Initiation

For a gene to be transcribed, the chromatin environment needs to be opened up for recruitment and action of the transcriptional machinery. Major TFs have the ability to bind nucleosomal DNA at regulatory sites throughout the genome and initiate conformational changes or even displace nucleosomes directly or indirectly through the recruitment of chromatin remodelling factors (Frietze and Farnham, 2011).

Transcriptional initiation at protein-coding genes takes place at the core promoter. One well-studied element of the latter is a short consensus sequence, the TATA box, usually located 25 base pairs upstream of the transcriptional start site (TSS) at the 5’ end of a gene. The TATA box is the site of the preinitiation complex formation. A generalised view of PIC assembly consists of the TATA-binding protein (TBP), which is a subunit of the general transcription factor II D (TFIID), binding to the 5’TATA(A/T)A(A/T)-3’sequence. This process then causes the recruitment of several more general transcription factors like TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH (Alberts 2002). In the next step, the assembled PIC provides the DNA-directed RNA polymerase II (Pol II), a multi-protein (RBPs) complex which catalyses the synthesis of messenger RNA (mRNA) from protein-coding DNA (Young, 1991).

Besides the PIC assembly, other factors such as activators and repressors modulate the transcription initiation rate by influencing recruitment, assembly, and accessibility. For instance, the initial binding of TBP to the promoter can be assisted by the SAGA complex which inhibits TPB from promiscuous binding to DNA (Papai et al., 2020). The latter is changed to a more open confirmation through recruitment of the histone acetyltransferases CBP/p300 and the SWI/SNF complex providing space for Pol II to bind by removal of histone PTMs and nucleosome remodelling, respectively (Vo and Goodman, 2001; Clapier et al., 2017). Further, the impact of TF

binding varies depending on the local context and presence of other accessory proteins (Reiter et al., 2017).

Mediators of TF effector activity constitute large multi-subunit protein complexes which contain domains for chromatin binding and remodelling to facilitate bridging of TFs directly to Pol II. The mediator complex is recruited by many TFs bound to cis-regulatory elements and thus, associates with a great amount of active genomic loci. The interactions between TFs, mediator, and the PIC lead to large-scale structural changes, i.e. the formation of enhancer-promoter DNA loops (Allen and Taatjes, 2015). The mediator complex also interacts with the super elongation complex (SEC) influencing the next step of transcription, the elongation of Pol II in the 5' to 3' direction of the gene.

1.1.3.2. Transcriptional Elongation

Following the formation of the pre-initiation complex, transcriptional elongation begins with promoter escape of Pol II. The carboxy-terminal domain (CTD) of RBP1, a subunit of Pol II, is phosphorylated on serine residues at position 5 and 7 (Ser5-P, Ser7-P). These phosphorylation marks are deposited by the TFIIH-associated cyclin-dependent kinase CDK7 and cause the disruption of Pol II's interaction with the mediator complex (Wong et al., 2014). The CTD generally plays an important role throughout the gene expression cycle as an interactive hub for transcriptional regulation and localisation of chromatin as well as mRNA processing factors (Heidemann et al., 2013). It contains 52 tandem repeats of a YSPTSPS amino acid sequence (Meinhart and Cramer, 2004; Phatnani and Greenleaf, 2006). The serine residues of these repeats are unphosphorylated during initiation, ensuring interactions with the Mediator complex (Wong et al., 2014; Tsai et al., 2017).

Messenger RNA synthesis starts until Pol II pauses within a region of 25 - 35 nucleotides downstream from the TSS. In this paused state, Pol II stays stably associated with the DNA and is only capable of resuming when further signals elicit the progression into productive elongation (Nechaev et al., 2010; Kwak et al., 2013). Promoter-proximal pausing is present on the majority of actively transcribed and developmentally regulated genes in metazoans. It serves as a checkpoint for timed gene activation to readily switch from long-lived pausing to undergoing productive elongation, in addition to aiding the establishment of permissive downstream chromatin. It adds an additional regulatory step in the transcription cycle that allows activators or repressors of transcription to exert combinatorial control of transcription levels (Adelman and Lis, 2012). Pausing is established through TFIID and a complex formed by the negative elongation factor NELF (NELF-A, NELF-B, NELF-C/NELF-D, and NELF-E) and DSIIF (DRB Sensitivity Inducing Factor) that remain bound to Pol II until its release (Wu et al., 2003; Vos et al., 2016; Fant et al., 2020).

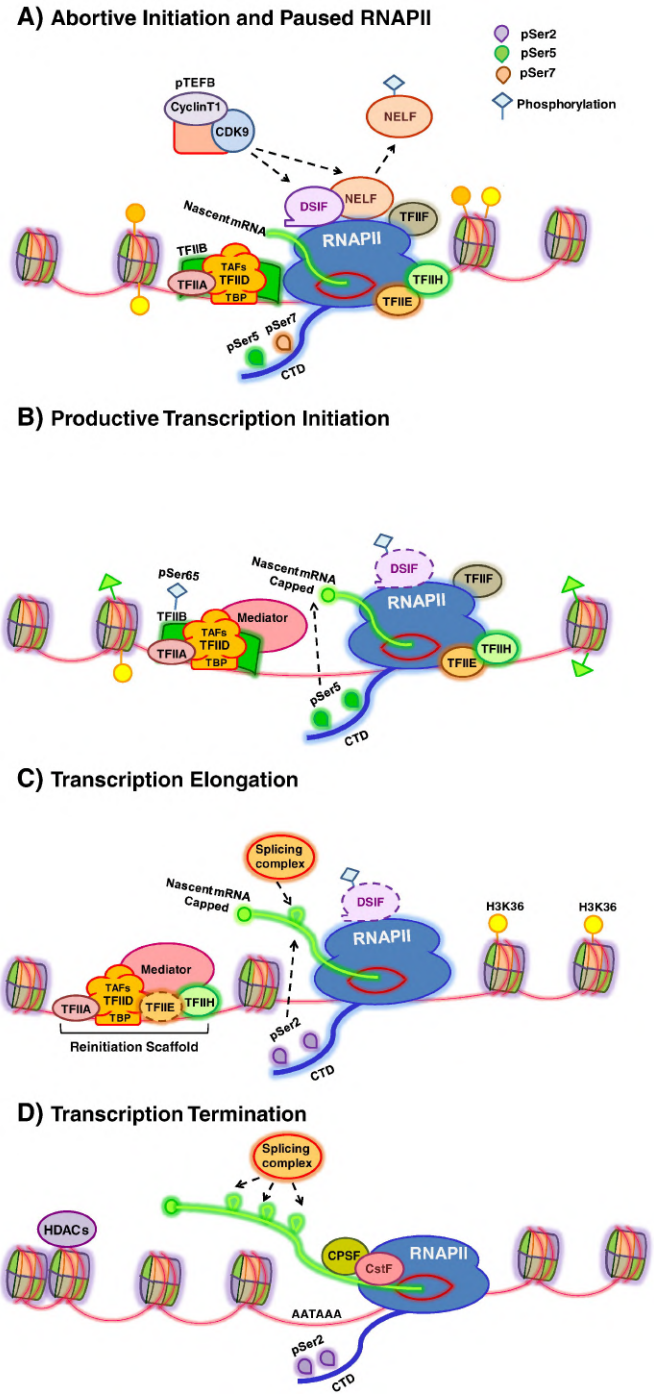


Figure 1 - The Process of Transcription.

Eukaryotic transcription is the process of converting a specific sequence of DNA into RNA in the nucleus. The main steps are **(A)** initiation, **(B)** pause release, **(C)** successful elongation, and **(D)** transcriptional termination. Adapted from Shandilya and Roberts, 2012.

Phosphorylation of the CTD residue serine at position 2 (Ser2-P) by CDK9, part of the positive transcription elongation factor b (P-TEFb), is essential to overcome the inhibitory effect of NELF. The paused Pol II is therefore released into productive transcription combined with gradual loss of Ser5-P (Peterlin and Price, 2006). Here, the activity towards serine 2 residue is only effective after phosphorylation of serine 7 by CDK7. Hence, successful transcriptional initiation is necessary for priming elongation (Czudnochowski et al., 2012). Towards the 3' end of the gene,

hyperphosphorylation on serine 2 is mediated by CDK12 (Peterlin and Price, 2006; Bartkowiak et al., 2010).

Correspondingly, chromatin-modifying and RNA processing complexes which are attracted by Ser2-P of the CTD enrich around two kilo bases downstream of the TSS to facilitate successful Pol II elongation and co-transcriptional processes (more detail in chapter 1.1.4. and 1.3.2.) (Bartkowiak and Greenleaf, 2011; Jonkers et al., 2014). The elongation factor P-TEFb interacts with the Mediator complex and necessary cofactors (Peterlin and Price, 2006). It is mostly associated with the larger super elongation complex (SEC) which comprises various factors that promote P-TEFb binding and support RNA synthesis (Luo et al., 2012). P-TEFb also phosphorylates Spt5 as a critical step for pause release that leads to dissociation of NELF which further enhances the rate of elongation (Yamada et al., 2006; Cheng and Price, 2007).

To catalyse mRNA synthesis, Pol II carries out most of the necessary function intrinsically, omitting accessory factors. It contains several subunits such as helicases for unwinding DNA and DNA binding proteins for the sliding movement over the gene. Together, these accessory subunits make elongation rates incredibly constant and precise; allowing for acceleration while transcribing through genes and slowing down at exons (Bushnell and Kornberg, 2003; Gnatt et al., 2001; Jonkers et al., 2014). Pol II incorporates nucleotides to the newly synthesised RNA chain by a Brownian-Ratchet mechanism facilitating a forward and backward movement along the DNA template (Bar-Nahum et al., 2005; Martinez-Rucobo and Cramer, 2013).

Successful elongation also requires constant nucleosome (dis-)assembly. Each nucleosome which is encountered by Pol II functions as an inherent barrier due to the compact wrapping of DNA around the nucleosome core. First, to overcome these boundaries, torsional stress of the elongating Pol II creates supercoils that destabilise consecutive nucleosomes (Teves et al., 2014). Second, a number of specialised elongation factors have evolved to modulate chromatin structure, such as polymerase-associated factor 1 (PAF1), SPT6 and Facilitates Chromatin Transcription (FACT) (Orphanides et al., 1998; Squazzo et al., 2002; Ardehali et al., 2009). The latter functions to destabilise the nucleosome core by removing a single H2A/H2B dimer ahead of Pol II and re-depositing it after Pol II passaging. The transcribing Pol II uses one single DNA strand to direct the next nucleotide to be added to the 3’ end of the growing mRNA. This occurs in a transcription bubble of about 25 exposed DNA basepairs (Cramer, 2004). Eight nucleotides of newly-synthesised RNA remain paired to the template DNA, whereas the rest of the RNA releases the template allowing the unwound DNA behind Pol II to be again compacted by FACT.

During elongation, Pol II transcribes from the 5’ to the 3’ end of the DNA template, resulting in the nascent RNA from the exit channel. Pol II is also able to backtrack one nucleotide in the reverse direction by extrusion of the transcript through the nucleotide entry channel (Nudler, 2012; Sheridan et al., 2019). This backward movement is stimulated by TFIIS and facilitates elongation as well as proofreading (error rate roughly 10^{-5}) of the newly introduced nucleotide during elongation (Blank et al., 1986; Sheridan et al., 2019). Pol II exhibits an intrinsic endonuclease activity cleaving the misincorporated nucleotide and allowing transcription to resume correctly (Sigurdsson et al., 2010).

1.1.3.3. Transcriptional Termination

Transcriptional termination at protein-coding genes results in the dissociation of the nascent transcript when the elongating Pol II passes the 3’ end of a gene. Pol II terminates transcription randomly with no specific location to stop and the newly-synthesised mRNA (pre-mRNA) is cleaved at its terminator sequence before transcription by Pol II ends (Porrua et al., 2016). The pre-mRNA’s cleavage site is found after an upstream AAUAAA sequence and before a GU-rich downstream part. If transcribed, these sequences are bound by cleavage-polyadenylation specificity factors (CPSFs) in a complex with WDR33 and FIP1 and by the cleavage stimulatory factor (CSTF), respectively (Nojima et al., 2015; Clerici et al., 2018). Additionally, the phosphorylation state of the CTD is of high importance for transcriptional termination: serine-2 is hyperphosphorylated (Ser2-P) towards the 3’ end and interacts with this cleavage complex, slowing down transcription (Hirose and Manley, 1998; Heidemann et al., 2013; Zhang et al., 2015a). The pre-mRNA is disengaged approximately 20 bp downstream of the AAUAAA polyadenylation site (PAS) and directly subjected to the polyadenylation machinery, catalysing the complementation of a poly(A)-tail to 3’ end of the mRNA (see chapter 1.1.5. for more detail) (Dreyfus and Regnier, 2002; Stewart, 2019).

It is generally acknowledged that Pol II continues transcribing past the GU-rich site of the DNA template. The biochemical details of the termination reaction after Pol II passes the PAS remain less understood; however, there are two models accepted as to how the termination process occurs: the torpedo and the allosteric model (Rosonina et al., 2006). The latter suggests that Pol II senses its passage through a PAS which induces a conformational change and its graduate dissociation from the DNA template prior cleavage of the nascent transcript (Osheim et al., 1999; Zhang et al., 2015a). Thus, termination is not completed by cleavage of the pre-mRNA, but is rather mediated by limiting the Pol II elongation efficiency, increasing the probability of ending the current transcription cycle.

In contrast, the torpedo model includes the requirement of cleavage of the nascent RNA. The remaining upstream part of the pre-mRNA is digested one nucleotide at a time by the 5'-3' exoribonuclease 2 (XRN2). XRN2 cleans up the residual RNA in kinetic competition ("torpedo") with the continuing Pol II. When reaching the latter, XRN2 is suggested to either push Pol II off the DNA by passing it or pulling the DNA out of the Pol II channel, both leading to transcriptional termination (Luo and Bentley, 2004; Luo et al., 2006).

1.1.4. Co-transcriptional Regulation

Eukaryotic pre-mRNAs are subject to extensive processing while being transcribed in the nucleus. RNA processing reactions are required to generate translatable, stable mRNAs, which are exported through the nuclear pore to sites of cytoplasmic translation (Bentley, 2014). Any mRNA that is not properly processed is degraded immediately after transcription (for detail see chapter 1.1.5.2.). It is critical to ensure this quality of the transcriptome which ultimately results in the proteome. In order to achieve such precise processing, many factors are recruited to the site of transcription by the CTD of the elongating Pol II and withdrawn in a time- and place-dependent manner, constituting a dynamic interaction network that governs the quality of mRNAs (McCracken et al., 1997; Ahn et al., 2004).

1.1.4.1. 5' Capping and 3' polyadenylation

Most eukaryotic genes exhibit a short RNA tract before and a longer one downstream of the coding region, the 5' and 3' untranslated region (UTRs), respectively. The 5' UTR starts with a 7-methylguanosine (m7G) cap, whereas the 3' UTR end exhibits a polyadenylated tail (Figure 2A and 2C). Both additions protect the mRNA from degradation and facilitate its export to the cytoplasm (Alberts, 2002). Additionally, these moieties serve as a recognition site for subsequent translational initiation and termination (see chapter 1.1.5.3.).

The 5' cap is introduced to the nascent mRNA as the DNA is still transcribed by the elongating Pol II. The terminal phosphate group at the 5' end is removed by the RNA triphosphatase, resulting in a bisphophate group (Wen et al., 1998). The mRNA guanylyltransferase utilises a GTP to form a 5'-5' triphosphate linkage to add guanine to the mRNA which is in turn methylated at position 7 by the mRNA guanine-N7-methyltransferase (Shatkin, 1976; Pillutla et al., 1998).

While Pol II is still transcribing downstream of the actual end of a gene, the pre-mRNA is cleaved between the AAUAAA polyadenylation consensus sequence (PAS) and a GU-rich sequence. The

functional pre-mRNA is released from Pol II and its 3’ end subjected to the process of polyadenylation. The mRNA 3’ processing is facilitated by a multi-protein polyadenylation complex which is attracted by the phosphorylated CTD of Pol II (Ser2-P) and signals transcriptional termination (Zhang et al., 2015a). This complex contains amongst others the cleavage/ polyadenylation specificity factor (CPSF) that catalyses the cleavage by recognising the PAS on the mRNA (Lutz, 2008). Specificity is increased through the cleavage stimulation factor (CstF) binding to the GU-rich site further downstream. The polyadenylate polymerase binds to CPSF and immediately adds about 250 adenosine monophosphates (poly(A)-tail) to the exposed 3’ end of the mRNA (Bienroth et al., 1993; Balbo and Bohm, 2007). The polyadenylation complex is also tightly associated with the spliceosome, a machinery that removes non-coding sequences from a pre-mRNA.

1.1.4.2. Splicing

The protein-coding DNA sequence is based on exons of a gene, which are separated through intervening intron sequences. These non-protein coding parts are precisely removed from the pre-mRNA and exons are rejoined in frame by a process called splicing, which results in correctly coding transcripts (Figure 2B) (Alberts, 2002). The responsible protein complex is the spliceosome which comprises subunits of small nuclear ribonucleoproteins (snRNPs), which are proteins coupled to the small nuclear RNAs (snRNAs, U1, U2, U4, U5, and U6), and additional complex members such as helicases (Wahl et al., 2009).

Another requirement for successful splicing is the interaction of the spliceosome with pre-mRNA. The introns of the latter contain signal sequences which include a 5’ splice site (starts with GU), a branch point sequence with a highly conserved adenosine residue, and a 3’ splice site (ends with AG). The excision of one intron is coordinated in two trans-esterification reactions on each 5’ and 3’ end. The spliceosome assembles on the pre-mRNA by interaction of U1 snRNA with the 5’ splice site, whereas the U2 snRNA binds to the branch point. This exposes the hydroxyl (-OH) group of the conserved adenosine residue as a nucleophile to attack the phosphodiester bond at the 5’ splice site. The remaining snRNPs assist the rearrangements and sequential release of the 5’ exon. Due to a conformational change in the snRNA U2 the interaction with the branching point is disrupted and U5 is able to interact with both the 5’ and 3’ splice site, bringing the terminal -OH group of the 5’ exon in close proximity to the last nucleotide of the 3’ splice site. Following a trans-esterification reaction at this site, the 3’ exon is connected to the 5’ exon, releasing the excised intron in a lariat form (Wahl et al., 2009; Chan et al., 2018).

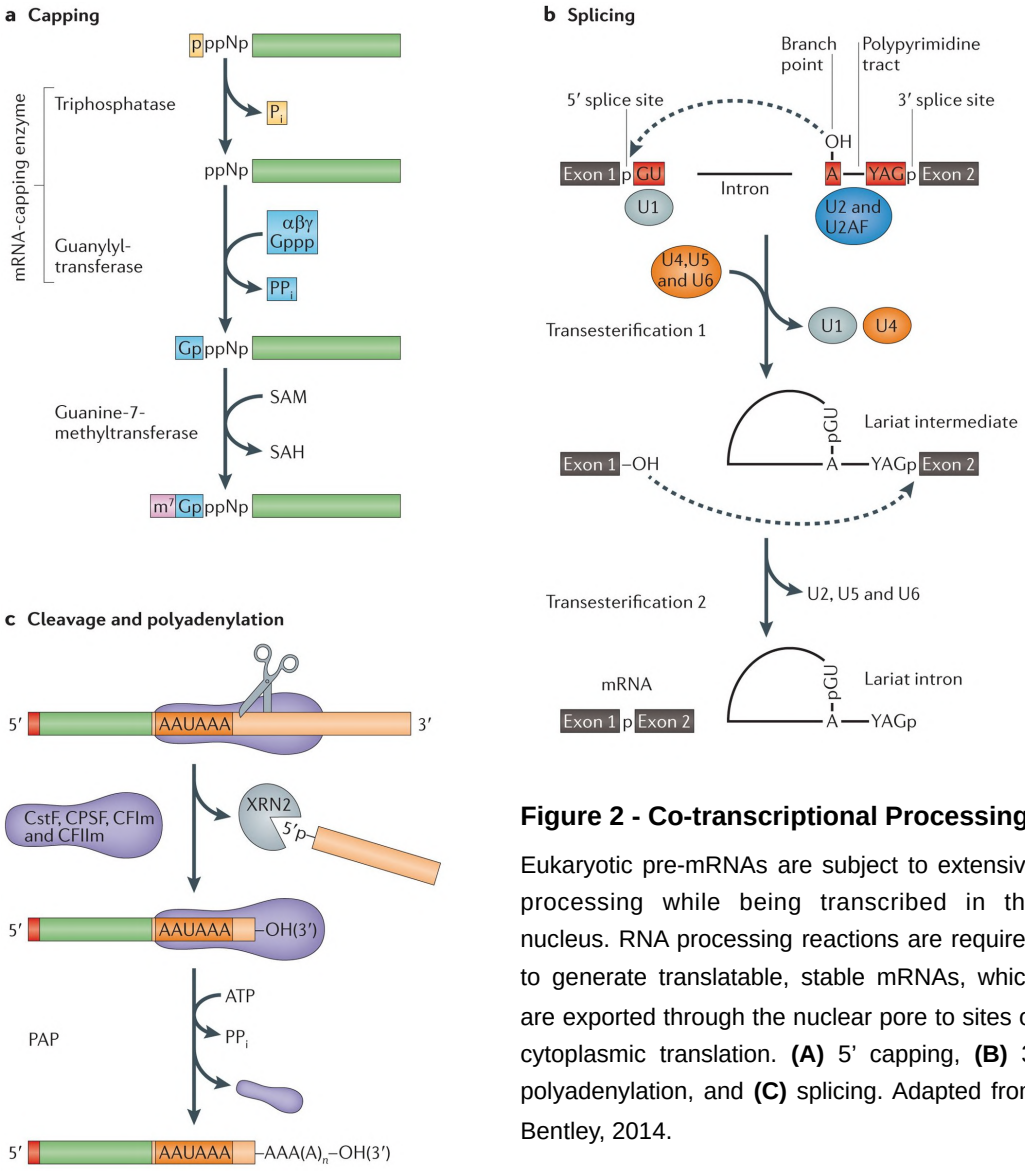


Figure 2 - Co-transcriptional Processing.

Eukaryotic pre-mRNAs are subject to extensive processing while being transcribed in the nucleus. RNA processing reactions are required to generate translatable, stable mRNAs, which are exported through the nuclear pore to sites of cytoplasmic translation. **(A)** 5' capping, **(B)** 3' polyadenylation, and **(C)** splicing. Adapted from Bentley, 2014.

All eukaryotes express mRNAs that are monocistronic. Once transcribed and processed, the mRNA can only be translated into one single protein. The process of alternative splicing adds a layer of multiplicity to this rule. Generally, the splicing occurs on every new arising splice site in a constitutive and orderly fashion during transcription. However, one pre-mRNA can give rise to several mature mRNAs for translation (isoforms) due to the use of alternative splice sites. This is facilitated by the exclusion of exons referred to as exon skipping, by alternative splice site usage which extends or shortens exons. Additionally, this process works through intron retention which includes an intron sequence in the mature transcript (Lee and Rio, 2015). Alternative splicing results ultimately in protein variants containing or lacking certain domains and thus, pursuing different functions in a tissue- and cell type-specific manner (Chen and Manley, 2009).

The pattern of alternative splicing further depends on cis-acting cues on the pre-mRNA, such as N6-methyladenosine, and on the context-dependent level and activity of regulatory splicing activators and repressors, like the heterogeneous nuclear ribonucleoproteins (hnRNPs) (Krecic and Swanson, 1999; Xiao et al., 2016; Zhou et al., 2019). Besides, components of the spliceosome are recruited and instructed by the phosphorylated CTD of the elongating Pol II (Phatnani and Greenleaf, 2006). The latter can interfere with the splice site interaction through its elongation rate. For instance, slower elongation leaves more time for the spliceosome assembly at weaker splicing sites, thereby excluding the adjacent (intron) sequence in the final transcript. Additionally, related chromatin structure and modifications influence splice site recognition and splicing machinery assembly, such as the histone modification H3K36me3 in gene bodies (see chapter 1.3.2.) or general nucleosome positioning during transcription (Moore and Proudfoot, 2009; Luco and Misteli, 2011).

There are also downstream consequences of alternative splicing, changing the abundance of mature mRNAs. For instance, sequence alterations can determine the efficiency of mRNA export from the nucleus or can alter the mRNA stability by constituting longer/shorter 3’ UTRs or containing pre-mature stop codons (see chapter 1.1.5.) (Moore and Proudfoot, 2009; Di Giammartino et al., 2011). Thus, splicing has a critical influence on the information transfer from transcription to translation and ultimately on the cell identity.

1.1.4.3. N6-methyladenosine (m6A)

Messenger RNAs also undergo dynamic and reversible chemical modifications to control gene expression. The most abundant modification on mRNAs is N6-methyladenosine (m6A) which is widely conserved among eukaryotic species and is considered to heavily influence mRNA life cycles (Roundtree et al., 2017; Zhao et al., 2017). It is observed in long exons, but mostly accumulates at the 3’ end of the transcript (stop codon, 3’ UTR) at the sequence motif DRACH (D = A, G, or U; R = G or A; H = A, C or U) (Dominissini et al., 2012).

m6A is set co-transcriptionally by a multicomponent methyltransferase complex containing the core enzymes Methyltransferase Like 3 and 14 (METTL3, METTL14). METTL3 exhibits the catalytic function to transfer a methyl group from the donor substrate S-adenosyl methionine (SAM) to the adenine nucleobases, whereas METTL14 lacks enzymatic activity, but stabilises the complex through structural support and binds to the RNA (Sledz and Jinek, 2016). Both methyltransferases associate with Wilms Tumor 1 Associated Protein (WTAP) which serves for RNA binding specificity and methylation efficiency (Ping et al., 2014). However, transcript specificity and choice of motif site for m6A remain only poorly understood.

Studies have shown that the m6A writer complex can be recruited to the mRNA in several ways: specific transcription factors at promoters or histone modifications, RNA-binding proteins as well as slowly elongating Pol II have been proposed to facilitate m6A (Slobodin et al., 2017). For instance, TGFβ signalling promotes methylation on a subset of mRNA in pluripotent cells (Bertero et al., 2018). Since m6A is enriched near stop codons, it is also likely that structural components on the mRNA, such as terminal exon-exon junctions constitute key features for its enrichment (Batista et al., 2014). Co-transcriptional enrichment for m6A also hints to an interaction with active Pol II and other chromatin features for transcriptional elongation. It was shown that METTL3 is directly recruited to Pol II and interference with Pol II speed correlates with m6A levels on the mRNA (Slobodin et al., 2017). Alternatively, exons show higher nucleosome density than introns and therefore enrich for active histone modifications, such as histone 3 trimethylation on lysine 36 (H3K36me3) (Kolasinska-Zwierz et al., 2009; Tilgner et al., 2009). Thus, enrichment for m6A on longer exons or towards the 3' end of the gene correlates with the accumulation of H3K36me3 on actively transcribed genes (see chapter 1.3.2.) (Dominissini et al., 2012; Huang et al., 2019). However, since there is no known domain of METTL3 or 14 for binding to H3K36me3 and loss of this mark only accounts for 40 % decrease of m6A, the specific mechanisms for transcript and site specificity required for m6A deposition are not completely elucidated, yet.

It is better understood which impacts m6A can have on mRNA properties and sequential processes. It directly effects the mRNA transcript through changing the charge, secondary structures, and interactions with proteins, which in turn shape transcript regulation by modulating mRNA processing, export, translation and decay (Roundtree et al., 2017; Zhao et al., 2017). YTH domain-containing proteins bind m6A and depending on their localisation play a role in these processes. For instance, YTH Domain Containing 1 (YTHDC1) supports exon inclusion in a small number of genes by recruiting or blocking different splicing factors (Xiao et al., 2016). Upon export to the cytoplasm, m6A is read by cytosolic readers like YTH domain-containing family protein 2 (YTHDF2 or DF2) which co-localises with both deadenylation and decapping enzyme complexes, promoting degradation of the destabilised mRNA (Wang et al., 2014; Du et al., 2016). In contrast, DF1 (YTHDF1) binds m6A and the eukaryotic translation initiation factor eIF3, a multi-protein complex that recruits the small ribosome subunit to mRNAs to promote their translation (Wang et al., 2015). Thus, m6A can generally be seen as a modification that decreases mRNA half-lives in the cell.

1.1.4.4. Regulation of Transcriptional Noise

Variability on gene expression between cell types in an isogenic population is referred to as transcriptional noise. It originates from fluctuations in the concentration, state, and location of shared molecular players in key processes such as mRNA expression or protein binding in a single cell at a given time. These fluctuations are caused in different ways and their degree varies between different genes and organisms (Balazsi et al., 2011). Gene expression noise is of stochastic nature and is distinguished in two types: extrinsic and intrinsic (Elowitz et al., 2002).

Extrinsic noise occurs gene-independently, carrying the potential to affect all genes in a similar manner. It originates from environmental perturbations in the surroundings of an individual cell, ultimately leading to local changes in the cellular concentrations of factors shared among genes like Pol II (Cao and Grima, 2020). Extrinsic features also include physical properties like cell size and growth rate. For instance, during mitosis cell contents separate unequally that can impact the availability of all molecular players and downstream processes (Huh and Paulsson, 2011).

In contrast, intrinsic noise is dependent on gene-to-gene variations from each cell to another. Thus, one gene can differ in its expression between two ore more cells due to fluctuation in the stochastic binding and diffusion of intracellular molecules. For instance, the activity of cis-regulatory elements (promoter, enhancers) depends on the presence and binding of TFs to these sites which is in turn based on the differential expression and stability of their mRNAs, resulting from a cascade of (co-)transcriptional processes including chromatin cues (Thattai and van Oudenaarden, 2001; Hebenstreit, 2013; Fujita et al., 2016).

Transcriptional bursting has been shown to be a major phenomenon that influences noise levels. It reflects fluctuations of promoters in their “on” or “off” state in irregular intervals leading to differential transcription rates at these genes. The frequency, length and strength of these bursts in a given period of time dictates the amount of mRNAs transcribed from the gene that can differ from cell to cell in a population (Kumar et al., 2015). Gene promoters which exhibit low nucleosome occupancy downstream of the TSS also show low transcriptional noise indicating that chromatin states and remodelling have a large influence on gene expression initiation (Cairns, 2009). Generally, transcription is a highly controlled process for which chromatin needs to be arranged in order to facilitate passaging of Pol II on nucleosome-depleted DNA. Chromatin accessibility is amongst others altered by the acetylation state of histone tails (see chapter 1.2.2.). Increasing acetylation levels downstream of the promoter results in open chromatin and correlates positively with bursting frequency (Struhl, 2007; Nicolas et al., 2018).

Each cycle of transcription, chromatin in gene bodies is opened and packaged again for successful Pol II elongation. This compaction is tightly controlled to prevent transcriptional initiation from sites other than the canonical 5' start of the gene which is referred to as spurious transcription or cryptic initiation. The latter can contribute to gene expression noise (Kaplan et al., 2003; Weinberger et al., 2012). Examples for these pervasive, non-coding transcripts are mRNAs from orphan promoters, lying within a gene, mobile repetitive elements (transposons) as well as antisense transcripts, since cryptic transcription may be bi-directional interfering with sense transcription (Jensen et al., 2013). Some of these RNAs may be functional, but mostly their uncontrolled transcription needs to be contained to avoid the disruption of cellular events like the production of the corresponding, canonical full-length transcript (Jacquier, 2009). Thus, there are lots of factors involved in the control of transcriptional directionality and chromatin accessibility at coding sequences. For instance, the SAGA complex with its histone acetyltransferase (HAT) and a histone deubiquitinase module is known to stimulate correct transcription initiation by guiding TBP to the promoter DNA. This in turn provides the correct directionality and choice of 5' start site for PIC assembly and transcription (Papai et al., 2020).

Further, the histone deacetylase Rpd3S recognises H3K36me3 in gene bodies and deacetylates these histones within transcribed regions after Pol II passage, thereby erasing the transcriptional memory and reestablishing the chromatin to a hypoacetylated state that is repressive for spurious transcription (see details chapter 1.3.2.) (Carrozza et al., 2005). The same holds true for SPT6 and FACT - histone chaperones which evict and reassemble nucleosomes in the wake of transcription in conjunction with other remodelling complexes to prevent spurious transcription (Kaplan et al., 2003; Mason and Struhl, 2003). Overall, these different parameters in and around the process of transcription ultimately determine the levels of mRNA for each gene in a cell and thus the amount of noise within a population.

1.1.4.5. Co-transcriptional DNA Repair

The integrity of the genome and the accuracy of transcription as well replication are prerequisites for cell survival. However, the cell frequently encounters DNA lesions from both environmental and endogenous sources that can cause DNA to undergo various modifications, including single or double strand breaks, base damage, helix distortions or inter-strand crosslinking. Therefore, the DNA damage response (DDR) has emerged as a checkpoint system that arrests the ongoing cell cycle through checkpoint kinases (CHK1/2) and as a cellular network to inspect the genome and repair these DNA lesions. Given the differences in these aberrant DNA structures, the DDR comprises several specialised pathways which influence cellular processes like transcription or replication to ensure genome stability (Jackson and Bartek, 2009).

Genes that are actively transcribed are endogenous DNA-damage hotspots, but they are also sensed and repaired in very fast manner (Ljungman and Lane, 2004; Fousteri and Mullenders, 2008). A specific pathway evolved in which single-strand lesions that block Pol II elongation are resolved - the well conserved, transcription-coupled nucleotide excision repair (TC-NER) (Brueckner and Cramer, 2007; Hanawalt and Spivak, 2008). The TC-NER is a sub-pathway of the ubiquitous process of nucleotide excision repair (NER) which removes lesions from the template DNA strands throughout the genome. Generally, this pathway senses bulky single-strand breaks which alter the DNA helical structure, e.g. lesions triggered by ultraviolet (UV) light, and disrupt DNA base pairing. Therein, the damaged nucleotide is first recognised, followed by the removal of a single-stranded segment around the lesion by an exonuclease. The missing DNA strand is resynthesised by a DNA polymerase using the sequence of the complementary strand as a repair template (Scharer, 2013). Further, transcription, replication and recombination are error-prone processes: wrong incorporation of nucleotides or incorrect base pairing govern the need for proofreading. General errors such as mismatches, insertions or deletions are corrected by the mismatch repair pathway choice, using a mechanism similar to the NER (Li, 2008).

Of all different classes of DNA damage, double-strand breaks (DSBs) are the most malignant. If left unrepaired, they can trigger cell death and, if misrepaired, they can cause deletions or chromosomal translocations; early incidences in carcinogenesis (Jackson and Bartek, 2009; Surova and Zhivotovsky, 2013). DSBs can be directly induced by ionising radiation, but they also arise indirectly, e.g. from two closely located single-strand breaks or during the repair of other lesions. For instance, during S phase of the cell cycle, replication forks can collapse, when encountering single-strand breaks, base damage, or the repair machinery at a lesion (Hanawalt and Spivak, 2008). Slowing or stalling of replication is referred to as replication stress which can ultimately lead to fork collapse and in turn to more DNA damage (Zeman and Cimprich, 2014).

DSBs are resolved by sensors, signal transducers, and effectors. The MRE11–RAD50–NBS1 (MRN) complex senses DSBs and triggers the ATR and ataxia telangiectasia-mutated (ATM) kinases, key transducers for the phosphorylation of the histone variant H2.AX (serine 139 residue, γH2AX) (Falck et al., 2005). The spreading of γH2AX around the damage amplifies the recruitment of additional DDR factors like the mediator of DNA damage checkpoint protein 1 (MDC1), the p53-binding protein (53BP1) or BRCA1 which determine the subsequent steps for repair (Cortez et al., 1999; Stucki et al., 2005; Fradet-Turcotte et al., 2013). For instance, 53BP1 is antagonised by BRCA1 and removed from the DSB site to favour one of the two DSB repair downstream

pathways: homologous recombination (HR) over non-homologous end joining (NHEJ) (Escribano-Diaz et al., 2013; Hustedt and Durocher, 2016).

NHEJ aligns the broken DNA ends back together and is independent of all cell cycle phases (Chang et al., 2017). In detail, DNA break ends are stabilised and prevented from resection by a DNA-dependent protein (DP) kinase holoenzyme which further attracts the DNA ligase IV (LIG4), X-ray repair cross-complementing protein 4 (XRCC4) and the XRCC4-like factor (XLF) (Graham et al., 2016). They are key effectors of NHEJ by aligning and ligating DSB ends to resolve the DNA break in a rapid, but error-prone fashion (Jackson and Bartek, 2009). In contrast, during S/G2 phase the cell commits to homologous recombination for its repair pathway of choice due to the presence of a template sister chromatid (Krejci et al., 2012). After initiation of DNA end resection by MRE11 with its endo- and exonuclease activity, a 3'-tailed, single-stranded DNA (ssDNA) is exposed. This ssDNA is bound by RPA for protection from DNA degradation and from the formation of secondary structures (Jackson and Bartek, 2009). RAD51 and RAD52 outcompete RPA with the help of BRCA1- and BRCA2 to form a presynaptic filament on the DNA (Sleeth et al., 2007). The latter searches for a homologous sequence to invade for the presynaptic complex to rectify the lesion in an error-free manner (Wright et al., 2018).

DNA replication during S phase is initiated at numerous origins, whereas transcription takes place throughout the cell cycle. Thus, these two processes need to be facilitated in a coordinated manner to avoid replication stress (Wei et al., 1998; Hamperl and Cimprich, 2016). Although there are mechanisms in place to minimise spatiotemporal collision of the two processes, the interference of the transcriptional machinery with the replisome can occur in a direct and indirect fashion. Besides the actual physical collision of the two, their productive processing on the DNA can also be impaired through structural constraints of the DNA. Both processes require unwinding of the DNA, introducing torsional stress when the complex are in close proximity which can impede with their progression. Unresolved replication fork stalling ultimately leads to fork collapsing and the accumulation of replication stress-derived DNA damage (Zeman and Cimprich, 2014).

Recently, we showed that replication fork speed and origin firing can be further impaired by deregulation of the pre-mRNA cleavage machinery during transcriptional termination in mammalian cells. In particular, a multi-screening approach identified the pre-mRNA cleavage factor WDR33 to confer replication stress resilience through successful mRNA cleavage and release of Pol II at the TTS. Depletion of WDR33 thus resulted in increased replication stress and in the gating of nascent mRNAs together with their genomic origin to the nuclear periphery, undermining genome stability (see chapter 8.2. for details) (Teloni et al., 2019).

Chromatin cues are also important for successful DDR. Generally, gene expression regulation of signalling and repair factors, i.e. modulating their levels, can indirectly influence DDR pathways. In contrast, chromatin modifications and environment can also directly attract or repulse DNA repair proteins (D'Alessandro and d'Adda di Fagagna, 2017). For instance, chromatin-scaffold proteins are recruited to sites with DNA lesions, facilitating the formation of permissive chromatin for repair directly or indirectly through interaction with chromatin modifiers (Altmeyer et al., 2013; Wang et al., 2013). Additionally, DNA damage arising during active transcription is generally resolved quickly due to the presence of active chromatin marks such as H3K36me3 or acetylation for recruitment of specific DDR factors, fine-tuning the choice of DDR pathway (see chapter 1.1.4.5.) (Daugaard et al., 2012; Gong et al., 2015; Huang et al., 2018).

1.1.5. Post-transcriptional processes

In order to ensure that functional proteins are produced, gene transcripts are regulated at many different stages during and in between transcription and translation, the latter being the synthesis of proteins from mRNA.

1.1.5.1. Nuclear mRNA Transport

Translation takes place in the cytoplasm. Thus, all processed mRNAs have to be exported from the nucleus. This transport is facilitated by the highly conserved transcription-export complex (TREX) (Strasser et al., 2002) which contains the THO complex (THOC1, 2, 3, 5, 6 and 7) and additional factors such as RNA helicases ALY and UAP56. It binds to the 5’ region of the mRNA and recruits the export receptor NXF1:NXT1 heterodimer. This factor then locates to the nuclear pore complex (NPC) and promotes binding of TREX to the RNA. The latter changes its conformation which is required its active passage through the pore (Viphakone et al., 2019). NPCs are macromolecular assemblies that span the nuclear envelope and serve as barriers to the passive diffusion of molecules larger than 40 kDa. Thus, there is a need for cargo-carrier complexes to overcome this barrier in an energy-dependent manner. When reaching the cytoplasm, the transport system disassembles and is imported back to the nucleus (Stewart, 2019).

1.1.5.2. mRNA Degradation

mRNAs can be incorrect or exhibit improper function either because of mutations within their genes or because of errors made during transcription or co-transcriptional processing. It is particularly crucial to eliminate impaired mRNAs, since the resulting proteins can be non- or dysfunctional

having acquired dominant-negative or gain-of-function activity. Furthermore, controlling mRNA levels is a general cellular process to regulate the abundance of proteins and thereby the processes these proteins engage in. Thus, there are many quality control mechanisms in place to conduct mRNA surveillance at every stage of RNA biogenesis (Garneau et al., 2007).

As discussed in chapter 1.1.4., mRNAs are co-transcriptionally processed to increase their stability and enhance their translation. In particular, the 5' 7-methylguanosine cap and the 3' poly(A) tail are two integral stability determinants. Upon export to the cytoplasm, these structures interact with translation initiation complex elF4E and PABP, respectively, in order to protect the mature transcript from exonucleases. For mRNA degradation, either one of the two determinants need to be compromised; e.g. through decapping or the mRNA being internally cleaved (Garneau et al., 2007). Further, m6A methylation is known to regulate the stability of mRNAs. In the cytoplasm, the m6A-binding factor YTHDF2 co-localises with both deadenylation and decapping enzymes leading to a destabilisation and degradation of the mRNA (Du et al., 2016).

There are three different mRNA degradation pathways: Nonsense-mediated decay (NMD), no-go decay (NGD), and non-stop decay (NSD) (Garneau et al., 2007). NMD detects and degrades mRNAs with a premature stop codon which can originate from mutations and frame shifts by nucleotide deletions or insertions in the underlying gene. It is also activated by other features such as an unusually long 3′ UTR (Amrani et al., 2004). NMD is evolutionarily conserved and considered to be the most important RNA decay pathway since it acts before the energy-consuming translation of a corrupt mRNA is initiated. NMD is dependent on the activity of the RNA-dependent helicase and ATPase UPF1. The latter promotes the decay steps of NMD by translocating along single-stranded RNA in an ATP-dependent manner and facilitates the completion of mRNA degradation by unwinding the structured mRNA and abolishing the interaction of messenger ribonucleoproteins. Its activity is further stimulated by other NMD factors like UPF2 and UPF3X (Kurosaki et al., 2019).

In contrast, NGD and NSD act during translation of the mRNA at the ribosome. NGD recognises incorrect mRNAs through detection of stalled ribosomes, e.g. due to the presence of stable intramolecular or intermolecular mRNA secondary structures, truncations, or chemically damaged sequences. NGD endonucleolytically cleaves the mRNA near the stall site which is degraded by the exosome in a 3' to 5' fashion and by XRN1 in a 5' to 3' fashion (Doma and Parker, 2006). NSD is able to detect mRNAs with a premature poly(A) tail or without a natural stop codon, generated by breakage or by the absence of an in-frame stop codon. This causes translation to proceed along the poly(A) tail and forces the ribosome to stay associated to the mRNA. Therefore, NSD also facilitates the release mechanism of the ribosome and degradation of the utilised mRNA upon

normal translational termination (van Hoof et al., 2002). This mRNA decay is facilitated by the SKI complex which load a 3’ to 5’ exonuclease on nonstop mRNAs, and by the specialised ribosome rescue complex DOM34/HBS1 that binds to the 60S ribosomal subunit and triggers the ribosome quality control (RQC) pathway (Brandman et al., 2012).

1.1.5.3. mRNA Translation

After export from the nucleus, correct mRNAs are subjected to translation into functional proteins at the ribosome. Translation is a highly controlled process, as the rate and efficiency of translation is coupled to the metabolic and proliferative state of a cell (Malys and McCarthy, 2011). All mRNAs are translated in the 5’ to 3’ direction into a polypeptide chain from the amino to the carboxy terminus. Starting from the start codon (AUG), every triplet of bases codes for one amino acid in the polypeptide chain (genetic code) (Alberts, 2002).

After dissociation from the transcriptional machinery, mRNAs have been extensively modified (see chapter 1.1.4.). The 3’ poly(A) tail in particular is bound by the poly(A)-binding protein (PABP) which promotes mRNA stability and export to the cytoplasm. There, PABP also binds the translation initiation factor eIF4G. The latter is part of the eIF4F (eIF3, eIF4A, eIF4G, eIF4E) complex of which eIF4E associates with the 5’ cap of the mRNA. This locks the transcript in a loop structure necessary for translation to commence at the ribosome. The ribosome consists of two subunits, the 40S and 60S, which come together upon translation. For initiation, the small subunit is locked to the 5’ end through association with by eIF3, promoted amongst others by the modification of m6A on the mature transcript (Wells et al., 1998; Wang et al., 2015).

Translation requires the interaction of three different kinds of RNAs: mRNA, transfer RNA (tRNA) and ribosomal RNA (rRNA) as well as various adaptor proteins. Upon binding of the ribosome to the mRNA, the 43S preinitiation complex examines the mRNA to identify the start codon AUG which codes the starting amino acid methionine. tRNAs are non-coding RNAs which deliver amino acids to the ribosome. They are charged with an amino acid that is covalently bonded by aminoacyl tRNA synthetases. For specificity, they also exhibit an anticodon site which is complementary to the mRNA codon. Translation elongation starts when the methionine-charged initiator tRNA associates with the ribosome by eIF2, leading to structural changes and the assembly of both subunits to the complete ribosome (Sonenberg and Hinnebusch, 2009).

During elongation of the nascent polypeptide chain, amino acid are added in three steps: first, each new tRNA is positioned to the A site of the ribosome. Second, the correct amino acid is covalently bound to the chain and third, the mRNA shifts by one codon towards the 3’ end. Translational

elongation is an energy-consuming process and is facilitated by elongation factors such as eEF2 which travel with the ribosome and utilise ATP (Schuller and Green, 2018). Further, eIF4A of the eIF4F complex is an ATP-dependent helicase that aids the ribosome by resolving mRNA secondary structures that could interfere with the translation process (Hellen and Sarnow, 2001).

Correct assignment of amino acids to the nascent peptide chain is crucial. Even though the ribosomes are usually considered accurate, it is subject to errors that can lead either to the synthesis of erroneous proteins or to the premature abandonment of translation. There are mechanisms in place which co-translationally sense aberrant peptide chain sequences. For instance, peptides, which are truncated due to immature STOP codons on a wrong mRNA template, are ubiquitinated and immediately degraded by the proteasome. This is usually signalled by a stalled ribosome to the NGD pathway which releases and degrades the defect mRNA from the ribosome (Joazeiro, 2019). The ribosomal subunits are disassembled and rescued in a process mediated by the DOM34/HBS1 complex. The complex association with the 60S subunit is subsequently recognised by the RQC pathway, leading to the rapid degradation of the immature peptide chain (Brandman et al., 2012).

The termination reaction of the translational process is dependent on the release factor eRF1 which induces the disassembly of the ribosome-mRNA units upon recognition of one of the STOP codons (UAG, UAA, UGA). Upon hydrolysis from the ribosome, the nascent polypeptide chain spontaneously adapts a well-defined three-dimensional structure, its biologically functional protein conformation, mainly guided by hydrophobic amino acid interactions as well as proteins called chaperones (Alberts, 2002).

Proteins also undergo post-translational modifications (PTMs) which diversify the limited pool of 20 amino acids exponentially. PTMs are pivotal for the regulation of protein stability, distribution in the compartments of the cell and their interaction as well as function. Sites on the protein that carry PTMs are often nucleophilic such as the hydroxyl groups found on serine, threonine, and tyrosine, or the amine groups of lysine, arginine and histidine. PTMs can additionally occur on the or N- or C-terminus of proteins themselves. Examples for PTMs are phosphorylation, methylation, acetylation or ADP-ribosylation which are catalysed by a large family of enzymes specific to each modification. They are also in many cases readily reversible due to the presence of deconjugating enzymes; allowing quick changes in interactions and function (Millar et al., 2019). Especially histone proteins are heavily post-translationally modified which influences chromatin structure and gene expression regulation in various way (further elaborated in chapter 1.2.2.) (Bannister and Kouzarides, 2011).

1.1.6. Regulating Gene Expression in a Cell Developmental Context

All cells contain the same full complement of DNA, but each cell type only accesses portions of the genome in a spatiotemporal fashion to express a gene program that results in a specific proteome which is relevant for the cell's lineage commitment and function. During development, cells undergo differentiation from a multi-potent precursor state to a specialised, mature cells in order to function in a network amongst other cell types, ultimately leading to the formation of independent tissues and functional organs. This is considered a tightly controlled, irreversible process, regulated by driving transcription and growth factors which guide the activation and repression of various genes. Thus, cell fate specification results from a framework of lineage-specific signalling cues (Goldberg et al., 2007).

Conrad Waddington first described this concept of an epigenetic landscape for the process of cellular decision-taking during development. Using a metaphor, the cell was represented as a ball that encounters hills and valleys while rolling down towards its specific final fate. For reaching the latter, the cell commits to permitted and directed trajectories when entering a specific valley, prohibiting the option to go back up the hill of a less differentiated state (Waddington, 1957).

There are three basic categories of cells that make up an adult organism: germ cells, somatic cells, and stem cells. Germ cells originate in the primitive streak to the developing gonads. They undergo meiosis and produce mature, haploid gametes (eggs or sperm) for reproduction. Somatic cells only divide by mitosis, are diploid and give rise to any other specialised cell of the body than a germ or stem cell. Stem cells display an undifferentiated cell that carries the potential to differentiate into one identical and one more specialised cell (asymmetric division) or can divide for self-renewal into two identical daughter cells (symmetric division). Stem cells generally remain undifferentiated due to environmental cues in a particular niche; hence, they differentiate only if they leave their niche and receive differential intrinsic and extrinsic signals.

There are two types of stem cells: adult stem cells are found in most tissues in an adult, multicellular organisms, acting as a repair system for the body by maintaining the normal turnover of regenerative organs, such as blood, skin, or intestinal tissues. They are considered multipotent as they only give rise to a limited number of cell types. In contrast, embryonic stem cells (ES cells) arise from the inner cell mass of a blastocyst, the pre-implantation embryo (Thomson et al., 1998). They are pluripotent and generate the primitive ectoderm, which ultimately differentiates during gastrulation into all derivatives of the three germ layer: endoderm, mesoderm, and ectoderm,

producing multipotent progenitor cells that then give rise to all specialised cells in the adult tissues. For instance, the ectoderm forms amongst others the epidermis and the nervous system, the mesoderm constitutes the hematopoietic system as well as cardiac and skeletal muscular tissue, and the endoderm forms the internal lining of organ tissues (Alberts, 2002).

ES cells also display an established model system in research. They can be maintained indefinite under specialised culture conditions, are amendable to many changes, and easily manipulated to investigate mechanisms during pluripotency and upon differentiation to mimic developmental processes. There are two distinct states of pluripotency regulated by different factors and pathways in a murine in vitro system: 2i- and serum-cultured mES cells represent a naive or primed pluripotent state, respectively (Marks et al., 2012). The conventional ES cells are prone to differentiate - also referred to as being in a metastable pluripotent state - which is shown amongst others in the upregulation of genes which are already specific for lineage differentiation. As they are exposed to signals for both differentiation and self-renewal through serum and leukemia-inhibitory factor (LIF)-containing ES cell medium, these cells more closely represent the day E6.5 of murine embryogenesis. The earlier state of pluripotency (E4.5) can be received by reprogramming the conventional mES cells under 2i conditions using of two small-molecule kinase inhibitors (PD0325901, CHIR99021) (Ying et al., 2008; Ficz et al., 2013). They do not only block the differentiation process via the MAP kinase pathway targeting MEK, but also enhance self-renewal of mES cells by inhibiting glycogen synthase kinase-3 and activating WNT signalling. 2i-mES cells are postulated to represent the ground state of pluripotency. They are more homogeneous in morphology, exhibit a redistribution of trimethylation of histone H3 on lysine 27 (H3K27me3) and genome-wide DNA methylation is nearly absent, mimicking reduced lineage priming. On the contrary, the latter alterations are already established in the conventional mES cells, reflecting a more primed state when cultured in serum+LIF medium (Marks et al., 2012).

1.2. Chromatin Modifications

In response to cell-intrinsic and cell-extrinsic signals, precise control of gene expression is critical for maintaining cellular identity. The dynamic regulation of gene expression in eukaryotes depends on a complex cooperation between transcription factors, RNA Polymerase II and coregulators that amongst others modulate the chromatin landscape. Chromatin remodelling facilitates specific gene expression pattern which is established by changes in chromatin marks such as DNA methylation (5mC) and post-translational modifications of histones (PTMs) (Mikkelsen et al., 2007). Covalent modifications of cytosines and of histone tails constitute molecular marks which are written, read, and erased to distinguish between active and inactive chromatin. This influence on gene expression programs can be direct through altering chromatin accessibility or indirect by downstream effects dependent on the binding and activity of reader-domain carrying proteins.

1.2.1. DNA Methylation

Each cell within a multicellular organism has distinguishable properties established by its unique gene program. This specific cellular identity is determined by the expression of genes in a time- and place-dependent manner and is passed onto daughter cells by DNA sequence-dependent and -independent mechanisms. Heritable processes, which do not alter the DNA sequence itself, are termed “epigenetic” (Bird, 2007).

Methylation of cytosine bases is one of the best mechanistically understood epigenetic modification and plays various roles in genome regulation. Three conserved enzymes, DNA methyltransferase 1 (DNMT1), DNMT3A, and DNMT3B, are responsible for the deposition of methylated cytosine bases in mammals (Smith and Meissner, 2013). Therein, DNMT1 is considered the maintenance enzyme which restores the fully methylated state of the DNA after replication, whereas the DNA methyltransferases 3A and 3B catalyse *de novo* DNA methylation. All three enzymes are essential for normal development (Okano et al., 1999). Individual deletions of *Dnmt1* and *3b* as well as the combined knock-out of *Dnmt3a* and *3b* induce embryonic lethality in mice (Lei et al., 1996).

1.2.1.1. Functions of DNA methylation

DNA methylation is the most common chemical modification of cytosines, is accurately propagated during cell division and associated with long-term repression of transcription (Holliday and Pugh, 1975). In mammals, DNA methylation occurs at the majority of CpG dinucleotides throughout the

entire genome, except for those in regions of CpG-dense, but unmethylated DNA, referred to as CpG islands (Zemach et al., 2010). The latter are associated with promoters of actively transcribed genes. Methylation of such CpG islands leads to stable heritable transcriptional silencing (Weber et al., 2005). Further, DNA methylation is involved in the repression of entire chromosomes during development (mammalian X inactivation) (Mohandas et al., 1981), is required for the silencing of transposable elements (De La Fuente et al., 2006), and the regulation of genomic imprinting (Rainier and Feinberg, 1994). It has also been associated to processes involved in ageing and tumorigenesis (Egger et al., 2004; Jones et al., 2015).

Given the dynamic distribution and effects of this epigenetic mark, it has been well studied in many different mammalian cell lines and tissues at various stages of development. To gain a detailed overview as to how recent observations have helped to understand the context-dependent roles of DNA methylation, we put together a review publication which can be found in chapter 7.1. (Ambrosi et al., 2017).

We further described a ChIP-based method to investigate the binding preferences of mammalian DNMTs, using prior biotin-tagging of the latter. Therein, the ease of biotin-avidin interactions circumvents limitations arising from low-specificity, commercially available antibodies and ensures reproducible results in a high-throughput fashion (see chapter 7.2. for more details on this book chapter) (Manzo et al., 2018).

1.2.1.2. Genic DNA methylation

Unlike its role at promoters, DNA methylation of gene bodies is compatible with active transcription of the underlying gene and even scales with transcriptional activity. In mES cells, it is preferentially established by the *de novo*-methyltransferase DNMT3B (Baubec et al., 2015). In spite of the biological relevance of this enzyme and genic DNA methylation for development, it is not fully understood to what extent they cause further transcriptional regulation.

A precondition to address these questions is comparing the distribution of cytosine methylation throughout the genomes of a variety of species. Given the hyper-mutability of methylated cytosines to thymine, methylated CpGs are unexpectedly enriched within gene bodies (Bird, 1980). Conservation, but also differences between epigenomes within and between eukaryotic groups raises the question whether there is a common underlying mechanism at work, or whether the gene body DNA methylation has evolved for distinct biological roles in different organismal groups (Suzuki and Bird, 2008; Zemach et al., 2010).

One suggested function, which retains the concept that DNA methylation is a transcriptional repressor, could be the reduction of transcriptional noise (Suzuki and Bird, 2008). It has already been observed that intragenic DNA methylation is negatively correlated to transcriptional interference owing to spurious initiation within an active transcription unit (Huh et al., 2013). This idea can be extended by the fact that DNA methylation was primarily a mechanism for silencing of repetitive DNA elements, such as retroviruses or transposable elements. Thus, genic DNA methylation can block the initiation of transcription at these elements allowing at the same time error-free transcription of the host gene.

In general, complex transcriptional initiation at sites other than the 5' end is a common phenomenon in Drosophila and humans and has to be strongly regulated (Suzuki and Bird, 2008). For instance, cryptic transcription has to be prevented from "orphan promoters". These promoters might have been used at early stages of development, but escaped their repression by DNA methylation in the germline having maintained their high CpG density (Illingworth et al., 2010). It has also been indicated that DNMT3B-mediated *de novo* DNA methylation in gene bodies prevents a random entry of Pol II and subsequent spurious initiation of transcription in mES cells (Neri et al., 2017). However, the observed loss of intragenic methylation was not complete (max. 10-20 % reduction), while complete removal of methylation as observed in *Dnmt*-triple-KO cells or mES cells grown in 2i did not lead to increased cryptic transcription (Teissandier and Bourc'his, 2017).

Another co-transcriptional mechanism which might be influenced by gene body DNA methylation is splicing (see chapter 1.1.4.2.). It has been indicated that exons are often more highly methylated than introns, and transitions in the level of DNA methylation often occur at exon-intron boundaries (Maunakea et al., 2010; Shukla et al., 2011). Pol II has been shown to accumulate preferentially at spliced exons, leading to the hypothesis that DNA methylation dynamics can lead to the reduction of Pol II elongation rates which facilitates splice site recognition and spliceosome assembly (Jonkers et al., 2014). However, numerous questions are left unanswered about the direct role of genic DNA methylation in transcription and splicing.

1.2.2. Post-Translational Modifications of Histones (PTMs)

Proteins can be chemically modified after translation, e.g. with the addition of methyl, phosphate, acetyl or ubiquitin groups. The presence or absence of these marks regulate the protein activity, stability, and localisation. For instance, phosphorylation can result in either activation or deactivation of a factor in a signalling cascade, addition of ubiquitin can flag proteins for their

degradation by the proteasome, or most commonly the addition of methyl groups can result in protein-protein interactions that serve transcriptional regulation, stress response, or nuclear transport.

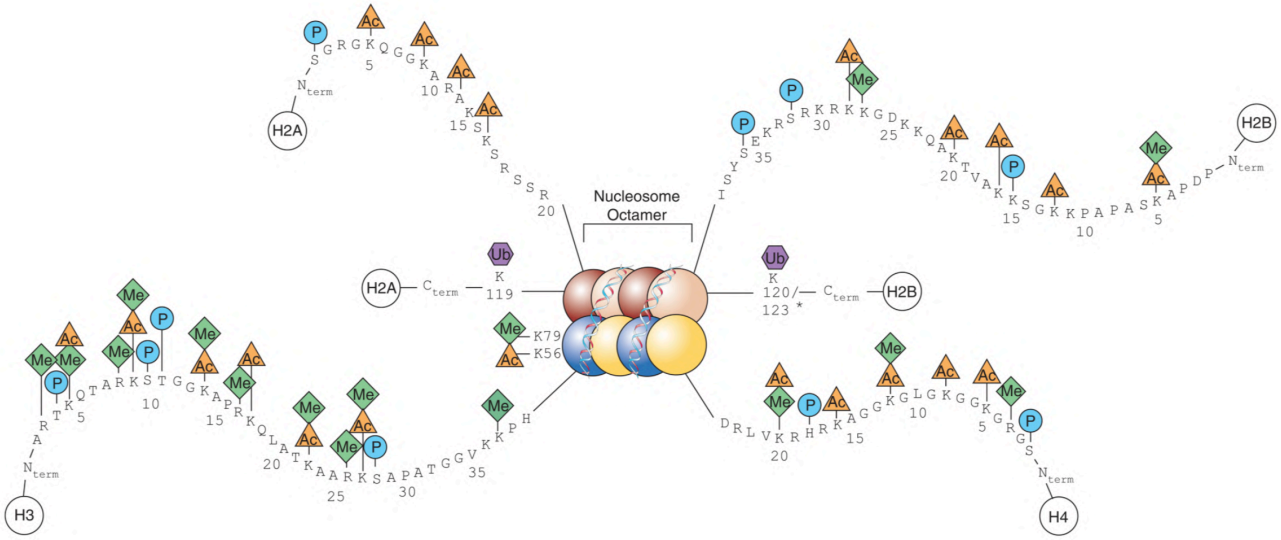


Figure 3 - Sites of important post-translational modifications on the histone tail.

PTMs found on the tails of the core histones H2A, H2B, H3, and H4 include Me - methylation, Ac - acetylation, P - phosphorylation, and Ub - ubiquitination. From cusabio.com

Particularly, covalent post-translational modifications (PTMs) of the core histones aid such diverse DNA-related processes influencing transcriptional activity, chromatin packaging and DNA damage response. PTMs are universal regulatory components among eukaryotes from yeast to human and act by directly determining the level of DNA compaction and indirectly by influencing the recruitment of regulatory effector proteins such as transcription factors, chromatin-modifying enzymes (chaperones) and effector proteins. Such actions result in diverse consequences for gene expression (Li, 2002).

Histones are most often modified on their N-terminal ‘tail’ region which is exposed and therefore easily accessible (Figure 3) (Luger et al., 1997). As well as the tails, other histone regions can also be modified such as the outside-facing region of the octamer, referred to as the lateral surface. This external globular histone domains can be extensively modified and have direct contact with the DNA, resulting in an important role in all DNA-based processes. However, these globular modifications are less studied, but not focus of this thesis work. They can be further reviewed in (Lawrence et al., 2016).

The histone tails are 40 amino acid long protein domains which are rich in basic amino acids, such as lysines and arginines (Figure 3). Given the vast number of different histone tail marks, the combinatorial possibilities are extensive which greatly multiplies the information content encoded

by the molecule. The levels of histone modifications are accurately balanced throughout the genome with distinct distributions and roles in the transcriptional state. Histone modifications are established through a dynamic cooperative action between histone PTM writers, erasers, and effector proteins called readers which bind depending on the neighbouring amino-acid sequence and the individual modification state. They can further be recruited by sequence-specific, DNA-binding transcription factors. Additionally, the act of transcription as well as general features such as the DNA CpG content and DNA methylation status can direct their recruitment (Li, 2002; Jambhekar et al., 2019).

Eukaryotic gene expression is greatly facilitated through the synergistic collaboration of transcription factors and the transcriptional machinery, chaperones and the presence of specific chromatin modifications. For instance, active enhancers are marked by the presence of H3K27 acetylation, active promoters are marked by H3K4 trimethylation, and active gene bodies are enriched with trimethylation on histone H3 Lysine 36. In contrast, PTMs coinciding with repressed genes encompass a broad domains of trimethylation of lysine 27, trimethylation of H3 lysine 9, and lysine 119 ubiquitination on Histone 2A (Jambhekar et al., 2019).

Of all the histone marks, histone lysine methylation is highly important and has been extensively studied. It is established by SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain-containing methyltransferases which bind and utilise the cofactor S-Adenosyl methionine (SAM) to add a methyl group (-CH_3) to lysine residue in an anabolic, nucleophilic reaction. This process converts SAM to S-Adenosyl homocysteine (SAH), which in several steps is regenerated back into SAM, referred to as SAM-e cycle (Trievel et al., 2002). The SET domain is evolutionarily highly conserved, carrying a 130-amino acid sequence motif representing the catalytic core. Its activity is further dependent on an adjacent pre-SET, cysteine-rich region which preserves the structure necessary for the catalytic reaction (Wood and Shilatifard, 2004). Methyltransferases often carry additional functional domains such as transcriptional activation/repression domains and domains for protein-protein or protein-DNA/RNA interaction to act in protein complexes and to mediate transcriptional output (Figure 4A) (Rea et al., 2000; Jambhekar et al., 2019).

Histone lysine methylation entails the addition of one, two or three methyl groups, adds hydrophobicity to the molecule, while only mildly changing the chromatin structure. Thus, its output is mainly facilitated by reader proteins which specifically recognise the methylation status with their chromo, PHD, PWWP, tudor, ADD, or ankyrin repeat domain (Figure 4B) (Hyun et al., 2017). The biological significance is also dependent on the combinatorial readout of multiple histone modifications, such as H3K4me3-H3K27me3 marked nucleosomes (bivalent promoters, see

1.2.3.1.) or the occurrence of phosphorylation on an arginine or threonine in close proximity to the methylated lysine (Fischle et al., 2005; Musselman et al., 2012). Therein, PTMs can cooperatively, antagonistically, or independently alter the local readout for transcription and chromatin compaction (Zhang et al., 2015b).

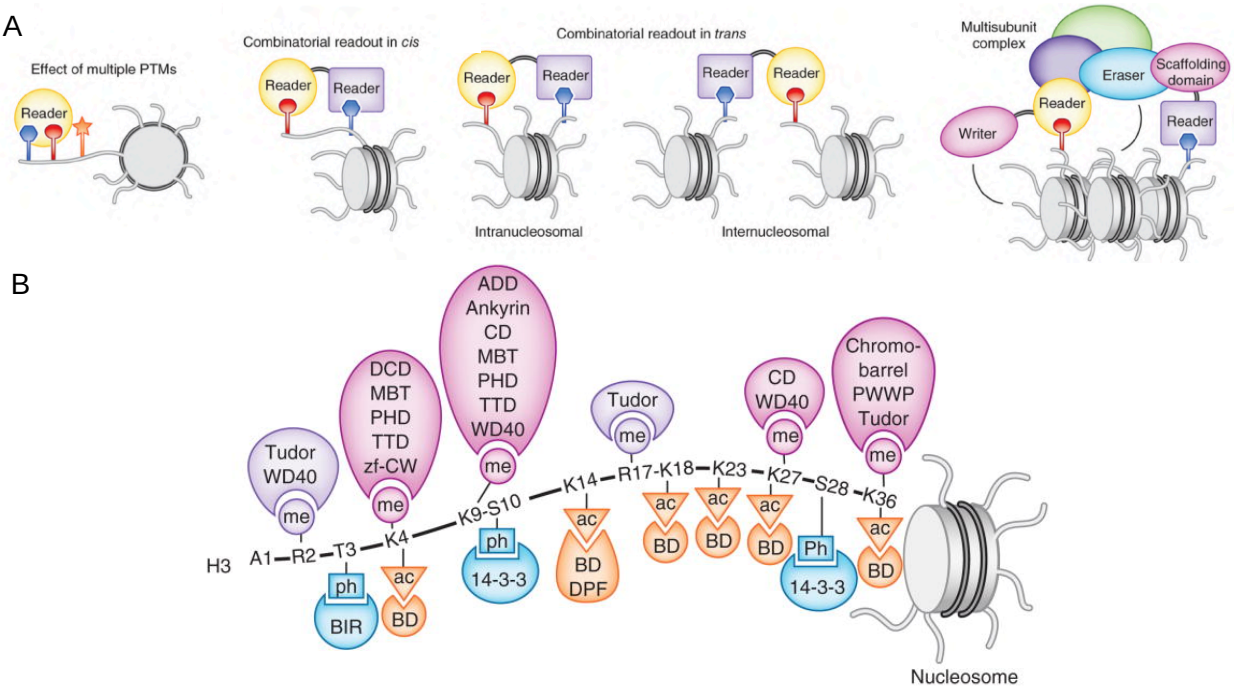


Figure 4 - Histone modification readout.

(A) Combinatorial readout of PTMs is facilitated through the influence by adjacent PTMs on the same or adjacent histone tail and through the action of multiple reader domains within the same protein or within subunits of an effector complex. **(B)** Recognition of histone PTMs found on the histone 3 N-terminal tail by different indicated reader domains (me - methylated, ac - acetylated lysine, ph - phosphorylated serine/threonine residues). WD - tryptophan-aspartic acid; PHD - plant homeodomain; ADD - ATRX, DNMT3, DNMT3L; BD - bromodomain; CD - chromodomain; PWWP - prolin-tryptophan-tryptophan-prolin. Adapted from Musselmann et al., 2012.

In order to study the biological readout of these reader domains, we have established ChromID, an experimental tool for analysing protein composition of specific chromatin environments (see chapter 8.1. for details). We used protein domains as interchangeable building blocks referred to as engineered chromatin readers (eCRs), binding to trimethylation of H3K4, H3K9 and H3K27 or DNA methylation to study their subnuclear localisation and genomic binding preferences in mES cells. In addition, we fused eCRs to a biotin ligase to unravel the protein network at bivalent promoters (see chapter 1.2.3.1. and 8.1.) (Villasenor et al., 2020).

1.2.3.1. Establishing Chromatin Environments

There are three main types of chromatin environments in the genome: active euchromatin, facultative and constitutive heterochromatin. A crucial function of these different chromatin states is to facilitate cell-type specific transcriptional activity and prevent the occurrence of genetic

instability. Each of these chromatin states is associated with a distinct set of chromatin modifications which orchestrate DNA-based biological tasks (Kouzarides, 2007). Euchromatin represents the largest proportion of the genome (92 %) (International Human Genome Sequencing, 2004). It is kept accessible for the regulation of gene transcription, for repair, replication or chromosome condensation and when actively transcribed, it exhibits a high degree of H3/H4 acetylation as well as H3K4, H3K36, and H3K79 trimethylation (Barski et al., 2007).

H3K4me3 is a hallmark for active euchromatin (Santos-Rosa et al., 2002). In mammals, six SET-domain containing methyltransferases (KMT2F/SETD1A, KMT2G/SETD1B, KMT2A/MLL1, KMT2B/MLL2, KMT2C/MLL3, KMT2D/MLL4) are responsible for this mark (Figure 5). Each one of them represents a catalytic subunit of large protein complexes by associating with four common core subunits: DPY30, WD repeat domain 5 (WDR5), retinoblastoma-binding protein 5 (RbBP5), and absent, small or homeotic-2 like (ASH2L). These writer complexes also contain accessory subunits that establish a distinct genomic binding behaviour (Hyun et al., 2017). The enzymatic redundancy is useful for targeting specific activities in a context-dependent fashion, such as methylation at an enhancer versus promoter regions, and for the selective establishment of the different methylation levels (me1, me2, me3) (Husmann and Gozani, 2019).

Generally, H3K4me3 levels coincide with CpG islands, the latter being associated with 50-70 % of promoters and appointing H3K4me3 as a mark for active transcription start sites (Santos-Rosa et al., 2002; Deaton and Bird, 2011). This is further supported by the association of SETD1 with the PAF complex and the initiating Pol II at promoters (Ng et al., 2003). H3K4me3-methyltransferases are also recruited by specific transcription factors or transcriptional coactivators engaging in the stimulation of gene expression (Zhang et al., 2015b). In contrast, H3K36me3 is located in gene bodies of actively transcribed genes due to its non-redundant writer SETD2, which associates with the Ser2-P elongating form of Pol II (Wagner and Carpenter, 2012). This particular histone modification is further elaborated in Chapter 1.3.

Besides H3K4me3, various H3/H4 acetylation marks coincide with cis-regulatory elements and transcriptional start sites of active genes. Histone acetylation is further spread throughout the body of actively transcribed genes and is heavily involved in transcriptional regulation and facilitation of co-transcriptional processes (Eberharter and Becker, 2002). There are many effector proteins which contain bromodomain to read acetylated lysine residues (Figure 4B) (Fujisawa and Filippakopoulos, 2017). H4 histone tail acetylation also directly effects DNA compaction by reducing the positive charge of histones. This abolishes electrostatic interactions with DNA,

resulting in a decreased chromatin compaction and increased DNA accessibility (Gansen et al., 2015).

Constitutive heterochromatin represents highly condensed, inactive domains in the genome which accumulate in the periphery of the nucleus by attachment to the inner nuclear membrane. It is composed of high copy number tandem repeats such as satellite repeats and transposable elements, mostly found at the pericentromeric or telomeric parts of chromosomes. However, it also occurs at perinuclear, euchromatic regions for the establishment of stable gene silencing which is critical for processes like differentiation and memory establishment (Zhang et al., 2015b).

The constitutive heterochromatic state is associated with low levels of acetylation and high levels of H3K9 as well as DNA methylation. The methyltransferases SUV39H1 and SUV39H2 catalyse H3K9 di- and trimethylation at satellite repeat regions in pericentric heterochromatin. Further, SETDB1 mediates silencing of specific target genes by H3K9me2/3 establishment through association with Krüppel-associated box (KRAB) associated protein 1 (KAP1) (Schultz et al., 2002). Similarly, the heterodimer of G9a and GLP targets H3K9me1/2 towards euchromatin in order to silence gene expression (Dillon et al., 2005; Tachibana et al., 2005). G9a and GLP both contain an ankyrin repeat domain which can recognise H3K9me1/2, allowing the enzymes to act in a feed-forward loop and spread the H3K9me2 modification (Figure 5) (Tachibana et al., 2002).

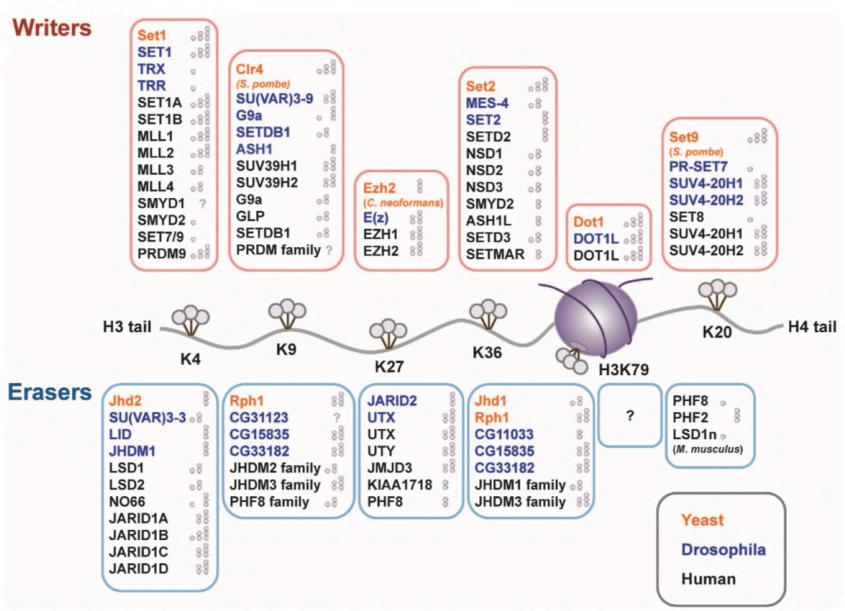


Figure 5 - Readers and Erasers of Histone Methylation.

Depicted are histone H3 and H4 methyltransferases and methylases with respect to different organisms and methylation levels (mono- (), di- (), and tri- () methylated. From Hyun et al., 2017.

This principle of propagation is further supported by the H3K9me2 and H3K9me3-dependent recruitment of the heterochromatin proteins 1 (HP1). HP1α and HP1β are co-recruited by SUV39H1/2 and bind to H3K9me3 with their chromodomains. This allows SUV39H1/2 to disseminate H3K9me3 to neighbouring nucleosomes (Bannister et al., 2001; Lachner et al., 2001).

Similarly, KRAB domain zinc finger proteins are the largest family of transcriptional regulators. They read H3K9me marks and recruit TRIM28/KAP1 which in turn attract HP1 and SETDB1, supporting the spreading of the repressive H3K9me3 histone mark especially on irreversibly silenced transposable elements (Ecco et al., 2017).

HP1 additionally recruits many other factors for heterochromatin formation like histone deacetylases, co-repressors and chromatin remodellers which leads to a tight compaction of nucleosomes and gene silencing (Hall et al., 2002). For instance, HP1 is known to recruit the *de novo* DNA methyltransferase DNMT3A. The resulting DNA methylation then reinforces the condensed chromatin structure (Fuks et al., 2003). Moreover, DNMT1, the proliferating cell nuclear antigen (PCNA) and G9a act synergistically during replication. They colocalise to the fork to set H3K9 methylation in order to faithfully maintain DNA methylation. This is further supported by the interaction between the ubiquitin-like, containing PHD and ring finger domain 1 (UHRF1) with H3K9 methylation, hemimethylated DNA and DNMT1 (see details in chapter 7.1.) (Liu et al., 2013).

In contrast, facultative heterochromatin represents genomic regions which harbour the opportunity to adopt open or closed conformations depending on temporal and spatial contexts. It is most often associated with genes of key transcriptional regulators important for biological processes such as genomic imprinting, X chromosome inactivation, or cell cycle control (Plath et al., 2003; Hidalgo et al., 2012). For instance, during early development the establishment of the different embryonic lineages is tightly regulated by expressing a proper dosage of gene products at a specific point in time. Here, genomic imprinting, which leads to expression of a gene from only one allele through the silencing of one of the parental copies, becomes crucial for this developmental stage and is tightly regulated through the introduction of facultative heterochromatin (Inoue et al., 2017).

The facultative chromatin state is characterised by levels of H3K27 mono-, di- and trimethylation and H2A lysine 119 monoubiquitination (H2AK119ub1). All three states of H3K27 methylation are established by the Polycomb repressive complex 2 (PRC2) to create and maintain repressive heterochromatin. This evolutionary conserved Polycomb group complex contains four core subunits: the enhancer-of-zest homolog EZH2/1, SUZ12, EED, and retinoblastoma-binding proteins P46 and 48 (RBAP46/48) (Schuettengruber et al., 2017). EZH2 carries the catalytic SET domain, but alone it has no enzymatic activity (Figure 5). Only its affiliation in the PRC2 complex with other subunits permits methylation of H3K27 (Pasini et al., 2004). EED and SUZ12 therein recognise H3K27me3 and the histone tail itself, respectively, and spread the H3K27me3-repressive mark to neighbouring nucleosomes in a feedforward fashion (Hansen et al., 2008). Additionally, accessory factors associate with PRC2 like the AE binding protein 2 (AEBP2), JARID2

and polycomb-like proteins (PCLs), which are thought to modulate PRC2 local targeting specificity and activity, ensuring the spreading and maintenance of H3K27me3 domains (Zhang et al., 2015b).

H3K27me3 is a well-studied and integral mark of facultative heterochromatin. It is crucial for the repression of developmental genes in place- and time-dependent fashion (Zhang et al., 2015b). Various mechanisms of PRC-mediated gene silencing are known, such as direct chromatin compaction, chromatin looping, prevention of H3K27ac deposition by excluding the histone acetyltransferases (HATs) p300 and the CREB-binding protein (CBP), or inhibition of engaged Pol II (Simon and Kingston, 2009; Pasini et al., 2010; Schuettengruber et al., 2017). However, H3K27 methylation is not the only instance for facultative heterochromatin as it can serve as a signal for PRC1-mediated chromatin compaction. The canonical PRC1 is an E3 ubiquitin ligase complex which introduces mono-ubiquitylation on lysine 119 of histone H2A through its catalytic subunit Ring Finger Protein 1A/B (RING1A/B). It also contains Chromobox protein homologs 2, 4, 6, 7, or 8 which read H3K27me3, one out of six PCGF proteins, and the polyhomeotic-like protein 1 (PHC1) (Gao et al., 2012). Although it is of importance for the silencing of a great number of developmental genes, such as the Hox gene clusters, during early development, there remains a debate on how both Polycomb repressive complexes are initially recruited and in which order they act on each of their target genes (Bracken et al., 2006).

It is well established that Polycomb group protein activity is directly regulated through the interplay with other chromatin marks. For instance, H3K27ac is known for its role in gene activation and stands in competition with H3K27me3 for the same lysine residue on histone 3 - resulting in a mutually exclusive relationship between these two marks. During ES cell differentiation, the clearance of H3K27ac by the Nucleosome Remodelling Deacetylase (NuRD) complex enables the attraction of PRC2, facilitating a switch from high levels of acetylation to methylation on H3 lysine 27 at promoters (Reynolds et al., 2012; Kim et al., 2015). On the contrary, the presence of H3K4me2/3 and H3K36me2/3 in *cis* inhibits PRC2 activity through inhibition of EZH2 (Schmitges et al., 2011; Streubel et al., 2018). However, if only one of the two histone H3 tails is modified with H3K4me3, PRC2 can catalyse H3K27me3 on the other, resulting in asymmetrically decorated nucleosomes (Voigt et al., 2012). Most often, these nucleosomes coincide with CpG islands in the genome, referred to as bivalent promoters; the latter regulating developmental genes in ES cells (Mikkelsen et al., 2007; Mohn et al., 2008). Due to an engaged, but paused Pol II, the poised chromatin state can promptly switch to a transcriptionally active (H3K4me3 only) or inactive (H3K27me3 only) status upon differentiation cues (Bernstein et al., 2006; Bracken et al., 2006; Mikkelsen et al., 2007).

Genes containing these two bivalent markers exhibit higher levels of gene expression noise when compared to completely active or Polycomb-repressed genes, suggesting that both gene activation and silencing account for the stochasticity of gene expression. Further, genes with higher noise are often situated close to other fully Polycomb-repressed genes, suggesting that H3K27me3-heterochromatin spreading further increases transcriptional noise (Faure et al., 2017; Kar et al., 2017). Thus, there need to be mechanisms in place to ensure proper gene expression of developmental genes by reducing H3K27me3-domain spreading. At bivalent promoters, this is contained by DNA methylation at CpG shores, supported by the evidence that loss of DNA methylation leads to a widening of H3K27me3 domains to those shores (Lynch et al., 2012; Reddington et al., 2013).

We found a dynamic enrichment of DNMT3A1, the longer isoform of DNMT3A, to be preferentially localised to these H3K27me3-positive CpG islands shores (see chapter 8.3.). We also observed that this preferential recruitment overlaps with elevated 5-hydroxymethylcytosine (5hmC) levels, hinting to a balance between DNMT3A1 and TET activity to fine-tune PRC2 domain boundaries. We also showed that DNMT3A1 localisation follows H3K27me3-redistribution during differentiation (Manzo et al., 2018). How H3K27me3 itself is needed for the recruitment of this DNMT3A isoform remains to be shown.

1.2.3.2. Core Histone variants

Aside from various histone modifications, another layer of regulating all DNA template-based reactions is represented by histone variants. In eukaryotes, histones form the building blocks of chromatin, the nucleosomes, and display one of the most conserved and abundant protein families. The core histones H2A, H2B, H3 and H4 are replication-coupled being expressed only in S phase of the cell cycle and providing a stable source for packaging of newly replicated DNA. They are encoded by multiple gene copies that are present in clusters on different chromosomes and they lack introns as well as typical polyadenylation at the 3’ end, but contain a consensus stem-loop structure downstream of the termination site (Williams and Marzluff, 1995; Buschbeck and Hake, 2017).

Aside from these canonical proteins, histone variants have evolutionary emerged with special properties and localised functions. They are encoded by single genes and contain polyA-tails as well as introns which are spliced during RNA processing, resulting in alternative splice isoforms (Figure 6). They are expressed at very low levels, but throughout the cell cycle which provides variant-specific transcription and deposition at any time. There are eight variants of H2A (H2A.X,

H2A.Z.1, H2A.Z.2.1, H2A.Z.2.2, H2A.Bbd, macroH2A1.1, macroH2A1.2, macroH2A2) and six variants of H3 (H3.3, histone H3-like centromeric protein A (CENP-A), H3.1T, H3.5, H3.X, H3.Y). There are no known variants for histone 4 (Buschbeck and Hake, 2017).

Variants do not only differ in their expression from the replication-coupled canonical histones, but also in their genomic distribution. Whereas the canonical histones are equally incorporated throughout the genome, nucleosomes containing variants show distinct localisations, suggesting specific roles. For instance, H3.3 coincides with promoters and gene bodies of actively transcribed genes, whereas CENP-A accumulates in centromeric regions (Chow et al., 2005; Wirbelauer et al., 2005; Black et al., 2007). For proper temporal and spatial deposition of these histone variants, there are specific chaperones and remodelling complexes in place, such as complex, chromatin assembly factor 1 (CAF1) for DNA synthesis-independent and -dependent H3.1 and H3.2 deposition or histone cell cycle regulation-defective homologue A (HIRA) and ATRX for H3.3 incorporation (discussed in Mattiroli et al., 2015).

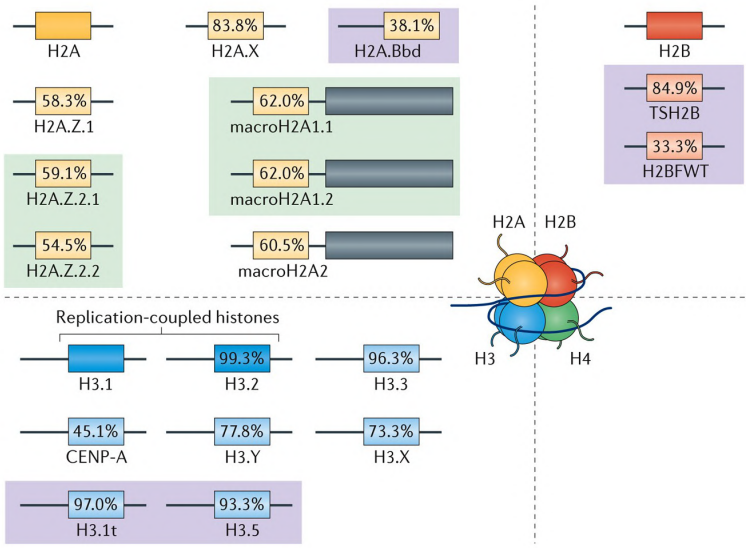


Figure 6 - Variants of Core Histones.
Depicted are variants of histone H2A, H2B and H3 (variants shown in pale colours). Rectangles represent core regions and lines represent flexible histone tails. Percentages indicate conservation (% sequence identity) of the variants relative to their replication-coupled counterparts. H2A.Bbd - H2A Barr body deficient; CENP-A - histone H3-like centromeric protein A; H2BFWT - histone H2B type WT; TSH2B - testis-specific histone H2B. Adapted from Buschbeck and Hake et al., 2017.

The replacement of core histones with variants increases the nucleosomal and functional diversity. This change in the nucleosome composition can alter chromatin structure in a direct manner and indirectly through specific reader proteins being sensitive to the histone variant status (Buschbeck and Hake, 2017). Further, due to their differential amino acid composition and genomic location, histone variant tails can be diversely modified. For instance, the phosphorylation of serine 139 of the H2A.X variant, referred to as γH2A.X, displays an early marker for DNA double strand breaks (Paull et al., 2000). It is read by the E3 ubiquitin-protein ligase RNF8 which facilitates chromatin decondensation through the interaction with the NuRD-complex member CHD4 (Mailand et al.,

2007; Luijsterburg et al., 2012). γH2A.X is further bound by mediator of DNA damage checkpoint protein 1 (MDC1), the MRN complex and ATM kinases which orchestrate the downstream DNA damage response (see chapter 1.1.4.5) (Turinetto and Giachino, 2015). Likewise, H3.3 often acquires modifications associated with active chromatin (H3K27Ac, trimethylation of H3K4, H3K36, and H3K79) and is deprived at silencing marks like methylation on K9 and K27 (McKittrick et al., 2004). The zinc finger MYND domain-containing protein 11 (ZMYND11) reads H3.3K36me3 in gene bodies of actively transcribed genes, linking histone variants to the modulation of transcriptional elongation (Wen et al., 2014).

1.3. SETD2 and Trimethylation of Lysine 36 on Histone 3

This study places emphasis on the elucidation of a functional role of genic lysine 36 trimethylation of histone 3 (H3K36me3) in transcriptional regulation. This histone modification is solely set by the methyltransferase SET domain containing 2 (SETD2) and decorates actively transcribed genes (Edmunds et al., 2008). The variety of reader proteins that interact with this mark suggest that H3K36me3 can directly as well as indirectly influence relevant pathways that control genome function and gene expression which remain to be elucidated.

1.3.1. Writers and Erasers of Histone 3 Lysine 36 Methylation

H3K36 methylation is an abundant and highly conserved mark in eukaryotes. In yeast, all three stages of H3K36 methylation are set by Set2, whereas in mammals eight different methyltransferases have evolved: Nuclear receptor-binding SET domain enzymes NSD1-3, SETD2 (KMT3A), SETD3, SETMAR, SMYD2 and ASH1L (Figure 5) (Wagner and Carpenter, 2012). SETD2 solely sets H3K36me3 with its conserved 130 amino-acid SET domain (Figure 7), restricting the other enzymes to mono- and/or di-methylation (further reviewed in McDaniel and Strahl, 2017). Interestingly, SETD2 does not require the preexistence of H3K36me2 to establish trimethylated K36 on histone 3 *in vitro*, as it exhibits higher affinity to unmethylated H3K36 (Sun et al., 2005; Kuo et al., 2011).

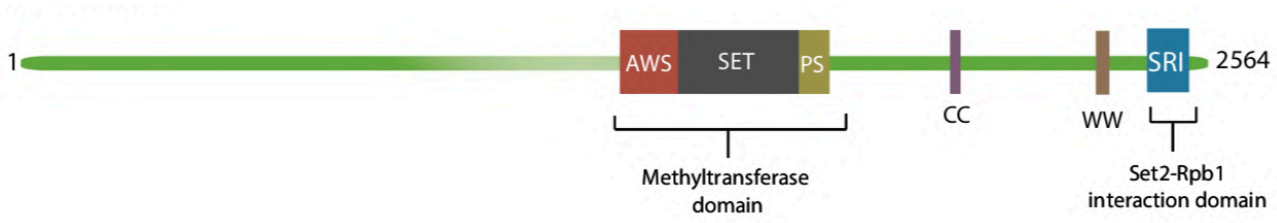


Figure 7 - Domain architecture of SETD2.

SETD2 exhibits a conserved domain organisation: Associated with SET (AWS), Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) and post-SET (PS) domains, all three required for the catalytic function; a coiled-coil (CC) and WW domain, and Set2–Rpb1 interaction domain (SRI), the latter binding to the Ser2-P CTD of Pol II. Adapted from McDaniel and Strahl, 2017.

There are distinct patterns of genomic distribution for the different methylation stages. Whereas H3K36me2 is more pervasive and enriched in both intergenic regions as well as proximal to the TSS, H3K36me3 is mutually exclusive to H3K4me3 and mainly found in the bodies (starting from around 2 kilobases into the gene) of actively transcribed genes (Santos-Rosa et al., 2002; Bannister et al., 2005; Edmunds et al., 2008). SETD2 binds with its SRI domain (Figure 7) the

phosphorylated Ser2-P C-terminal tail of the elongating Pol II (subunit RPB1) (Krogan et al., 2003; Edmunds et al., 2008). This interaction results in H3K36me3 spreading over the transcribed region, a mechanism which is conserved from yeast to humans. The uncoupling of H3K36me3 genomic distribution from H3K36me2 over evolution through different responsible enzymes indicates biologically distinct roles of the two methylation states (Wagner and Carpenter, 2012). This is further supported by the fact that NSD family members exhibit a PWWP domain that recognises H3K36me3, achieving a distinct genomic distribution (Vermeulen et al., 2010; Hyun et al., 2017).

In mammals, there are two families of H3K36 demethylases: JHDM3/JMJD2/KDM4A-D are responsible for demethylation of H3K36me2 and H3K36me3 (Hyun et al., 2017). In contrast, JHDM1/KDM2A-B only demethylates H3K36me1 and H3K36me2. This enzyme carries a PHD finger which recognises H3K4me3, explaining a mutually exclusive relationship between H3K4me3 and H3K36 methylation (Bannister et al., 2005). This notion is supported by the colocalisation of JHDM1 to H3K4me3-enriched CpG island promoters, ensuring active removal of H3K36me2 from the TSS (Blackledge et al., 2010).

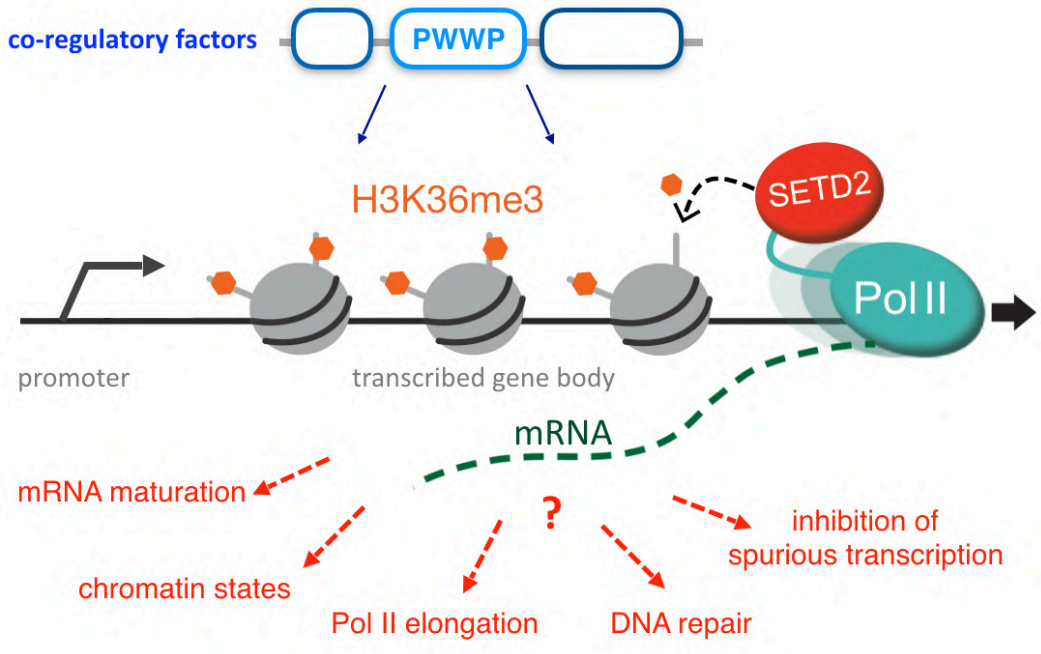


Figure 8 - Chromatin modifications on actively transcribed genes.

SETD2 interacts with the C-terminal of active RNA Polymerase II (Pol II) and mediates tri-methylation of lysine 36 on histone three (H3K36me3). Several factors are recruited by this mark via PWWP domains to actively transcribed regions, indicating a potential co-regulatory function. The role of H3K36me3 for mammalian gene regulation remains to be elucidated. Several mechanisms like RNA maturation or Pol II speed alteration could be influenced.

1.3.2. Functional Role of SETD2-mediated H3K36me3

SETD2 activity results in bookmarking of transcribed gene bodies by the H3K36me3 mark. SETD2/ H3K36me3 are essential for normal development since *Setd2* knock-out induces embryonic lethality in mice at day E10.5-11.5, with embryos showing growth and neurodevelopment defects (Hu et al., 2010). However, their role is not fully understood, but their association with active transcription and results from single-gene or correlative studies suggest that various regulatory mechanisms could be influenced (Figure 8).

In yeast, it has been shown that lack of histone H3K36me3 results in increased cryptic transcription (Sen et al., 2015). In particular, the H3K36me3 reader Rpd3S plays a pivotal role as a histone deacetylase in guaranteeing that gene bodies remain hypoacetylated after the Pol II elongates through. If its activity is impaired by loss of H3K36me3, increased acetylation can result in cryptic transcription initiation and altered deposition of histones across gene bodies (Li et al., 2003; Lee et al., 2013). Therein, Rco1 and Eaf3 components of the Rpd3S complex recognise H3K36me3, activating Rpd3S deacetylase activity. This suppresses antisense transcription from nucleosome-free regions as well as spurious cryptic transcripts from initiating within the open reading frame (Carrozza et al., 2005). H3K36me3 has also been shown to prevent the incorporation of acetylated histones (Venkatesh et al., 2012). Further, in higher eukaryotes, FACT complex subunit SPT16 binds to H3K36me3-carrying nucleosomes and promotes FACT-mediated exchange of histone H2B, but not H3, during transcription-coupled nucleosome displacement in order to maintain chromatin integrity during transcription (Carvalho et al., 2013).

In C. elegans, H3K36me3 is differentially distributed between exons and introns which is suggested to regulate exon selection for alternative splicing (Kolasinska-Zwierz et al., 2009). Based on this model, SETD2/H3K36me3 could cause severe effects on transcriptional elongation by directly or indirectly altering RNA polymerase II kinetics with influences on co-transcriptional splicing (de la Mata et al., 2003). This is further supported by MRG15 binding to H3K36me3, a protein that regulates alternative splicing through the attraction of the splicing factor polypyrimidine tract-binding protein (PTB) to the pre-mRNA (Luco et al., 2010). Reduction of SETD2 by proteosomal degradation is triggered via polyubiquitylation mediated by SPOP/CUL3 and induces alternative splicing events, supporting a role of H3K36me3 in splicing (Zhu et al., 2017). Further, ZMYND11 specifically recognises H3.3K36me3 and causes intron retention besides modulating the elongation rates of Pol II as a transcriptional co-repressor (Daugaard et al., 2012; Wen et al., 2014). It was also shown that H3K36me3 prevents intron retention potentially by marking weak splice donor sites and coordinates tissue-specific usage of alternative exons through the PWWP-

containing reader lens epithelium-derived growth factor (LEDGF) (Pradeepa et al., 2012; Liu et al., 2014).

The binding of LEDGF to H3K36me3 also links this mark to the DNA damage response system to maintain genome stability. LEDGF attracts carboxy-terminal binding protein and interacting protein (CtIP) to sites of DNA damage within transcribed genes. CtIP in turn facilitates binding of RAD51 which ultimately leads to the repair pathway choice of homologous recombination (HR) (Daugaard et al., 2012; Pfister et al., 2014). Generally, DNA repair of DSBs in actively transcribed genes is faster compared to inactive genes, suggesting a role for transcription in the repair of this type of lesions (Aymard et al., 2014). SETD2 binds to the transcribing Pol II and its loss ultimately results in the absence of H3K36me3 in gene bodies and decreased recruitment of important downstream DNA damage response factor such as 53BP1 or RPA (Aymard et al., 2014; Kanu et al., 2015). This in turn leads to reduced levels of ATM as well as p53 phosphorylation, impairing DNA repair by HR and threatens genome integrity (Carvalho et al., 2014). SETD2/H3K36me3 also act as regulators of DNA mismatch repair pathway in G1 and early S phase. H3K36me3 is bound by PWWP-containing MSH6 subunit of the MutS alpha complex to commence the mismatch repair reaction (Li et al., 2013; Huang et al., 2018).

Interestingly, similar roles in gene regulation have been proposed for gene body DNA methylation. It was shown that the *de novo* methyltransferase DNMT3B is particularly responsible for targeted DNA methylation of actively transcribed gene bodies in mES cells (Baubec et al., 2015). In relation to this, the presence of the lysine 36 trimethylation of the histone 3 promotes gene body DNA methylation by recruitment of DNMT3B through its PWWP domain, whereas H3K4me3 prevents the DNMT3B enzyme from adding methyl tags to DNA (Morselli et al., 2015; Rondelet et al., 2016). It had already been speculated before that genic DNA methylation is negatively correlated to transcriptional interference owing to spurious initiation within an active transcription unit (Huh et al., 2013; Neri et al., 2017). Another study showed that co-transcriptional splicing might also be influenced by gene body DNA methylation, again hinting to a potential interplay with SETD2/ H3K36me3 actions. This idea can be extended by the fact that exons are often more highly methylated than introns and transitions in the level of DNA methylation often occur at exon-intron boundaries (Maunakea et al., 2010; Shukla et al., 2011). Recently, a direct functional link between DNA methylation and SETD2/H3K36me3 was established in mouse oocytes where loss of SETD2 impaired the establishment of the correct DNA methylome. This resulted in aberrant genomic imprinting, oocyte maturation defects, and subsequent embryonic lethality after implantation (Xu et al., 2019).

This study in mouse oocytes also highlighted the importance of the interplay between histone marks. Loss of H3K36me3 in murine oocytes also led to the invasion and aberrant deposition of H3K4me3 and H3K27me3 into former DNA methylation and H3K36me3 territories (Xu et al., 2019). It is generally suggested that active marks such as H3K36me3 restrict the spreading of the silencing marks H3K27me3 through direct inhibition of PRC2 (Schmitges et al., 2011). This is amongst others facilitated by two PCL family proteins (PHF1/PCL1 and PHF19/PCL3) which constitute accessory components of the PRC2 complex and contain a Tudor domain with high affinity to H3K36 methylation (Brien et al., 2012; Cai et al., 2013). Indeed, it was recently found that H3K36me2/3 loss leads to genome-wide expansion of H3K27me3 at PRC2/PRC1 target genes (Streubel et al., 2018). Moreover, H3K27me1 correlates with H3K36me3 within transcribed genes, promoting active transcription, and its deposition is dependent on H3K36me3 in embryonic stem cells. SETD2 deficiency leads to an increase of H3K27me2 in gene bodies. The latter usually prevents firing of non-cell-type-specific enhancers and thus decreases gene expression when falsely deposited (Ferrari et al., 2014).

It is a well-established concept that histone methyltransferases can be enzymatically inactive or methylate non-histone targets. For instance, EZH2 was shown to not only trimethylate H3K27, but also the transcription factors STAT3 (Dasgupta et al., 2015). Similarly, it has been identified that SETD2 secures genomic stability by trimethylating lysine 40 of α-tubulin during mitosis and cytokinesis which are important processes in regards to the morphology changes during differentiation (Park et al., 2016a). Furthermore, SETD2 was shown to be involved in the interferon-alpha-induced antiviral defense by introducing monomethylation at lysine 525 of STAT1 (Chen et al., 2017). However, the recognition of such disparate substrates by SETD2 still require further investigation.

1.3.3. Clinical Relevance

SETD2 was suggested to function as a tumor suppressor in a highly tissue-specific context, suggesting a substantial need for SETD2 presence to avert tumorigenesis and thus could have important implications for future therapeutic strategies (Figure 9) (Duns et al., 2010). Downregulation of SETD2 is most often found in human clear cell renal cell carcinomas (ccRCC, 15-20 % of all cases) and is amongst others associated with aberrant RNA splicing (Dalgliesh et al., 2010; Kanu et al., 2015; Tiedemann et al., 2015). Therein, the pathogenic SET domain mutation (R1625C) as well as Pol II-interacting SRI domain mutation (R2510H) were found, suggesting individual roles for the enzymatic product H3K36me3 and the enzyme itself, respectively (Hacker et al., 2016)

Some tumors exhibit a more aggressive phenotype and worse prognosis as a pathogenic SETD2 depletion is considered to increase genome instability, promoting adaptability and clonal survival (Simon et al., 2014; Ho et al., 2016; Park et al., 2016a). After initial evidence for a potential tumor-suppressive role of SETD2 in ccRCC, subsequent sequencing studies have identified recurrent *SETD2* mutations across a broad spectrum of human malignancies, such as lung adenocarcinoma, bladder cancer or glioblastoma (Lawrence et al., 2014; Fahey and Davis, 2017; Husmann and Gozani, 2019; Chen et al., 2020). Furthermore, recent studies have identified mutations in histone 3-encoding genes at or near the lysine 36 residue, the SETD2 substrate, in a number of paediatric cancers including chondrosarcoma (K36M mutation) (Behjati et al., 2013; Fang et al., 2016) and glioblastoma multiforme (G34R/V mutation) (Schwartzentruber et al., 2012; Sturm et al., 2012), further highlighting the relevance of H3K36 methylation and the tumour suppressor SETD2.

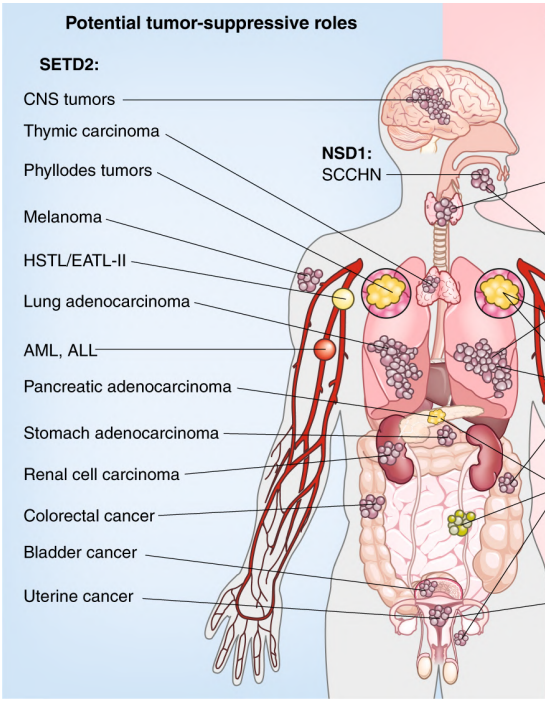


Figure 9 - Potential Tumor-Suppressive Roles of SETD2.
The potential tumor-suppressive functions of SETD2 by the identification of recurrent deletions, frameshifts, or truncating or damaging missense mutations, and by biological studies. CNS - central nervous system, HSTL - hepatosplenic T cell lymphoma, EATL-II - enteropathy-associated T cell lymphoma type II, ALL - acute lymphoblastic leukemia; AML - acute myeloid leukemia. Adapted from Husmann and Gozani, 2019.

In contrast, SETD2 loss can also support sensitivity of cancer cells to pharmacologic inhibition, suggesting a context-dependent role of SETD2/H3K36me3 in disease. This has been recently shown to be of high importance for acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) with MLL-fusion protein signature. Here, deficiency of SETD2 increased DNA damage in cells, reduced cell cycle progression, and promoted differentiation of AML cells (Skucha et al., 2018). Another study indicated that specific inhibition of PI3Kβ induces synthetic lethality with *SETD2* loss in ccRCC (Terzo et al., 2019).

As SETD2-mediated H3K36me3 is suggested to engage in the maintenance of transcriptional fidelity, its role could be also required to assure correct gene expression during the organism's whole lifespan by fostering longevity and could thus play a role in the process of ageing (Pu et al., 2015). Given these links to disease or ageing, understanding the role of H3K36me3 in transcriptional regulation will provide major contributions to human health. However, further work is necessary to elucidate the action of tumor suppression by SETD2, the relationship to H3K36me3 and the effects on downstream genetic context.

2. Aims

Although there is a clear correlation between chromatin modifications and gene expression states, it will be important to establish their direct role in regulating transcription and cell identity. In particular, given the crosstalk between SETD2-mediated H3K36me3 and other chromatin marks as well as chromatin-bound factors, the individual and overlapping roles in maintaining cell identity require to be explored in more detail.

Single-gene or correlative studies propose a complex picture of SETD2-mediated H3K36me3 function in various transcription-dependent and -independent processes. However, due to the lack of functional and comprehensive investigations in defined cellular models, the direct function of this mark and its enzyme in mammalian cells still remains be elucidated. With this thesis work, I wanted to address the role of SETD2 and H3K36me3 in regulating gene expression through a combination of genome-wide transcriptional readout measurements and functional rescue experiments in a highly-controlled cellular differentiation model.

I was able to address the following questions:

- **What is the contribution of chromatin modifications to gene expression noise and cellular identity of murine embryonic stem cells?,**
- **How do chromatin modifications regulate gene transcription during development?,**
- **What is the functional role of SETD2-mediated H3K36me3 for establishment and maintenance of gene expression programs?**

3. Exploring the Functional Role of SETD2/ H3K36me3 during Cellular Differentiation

3.1. Chromatin modifications do not influence gene expression noise in murine embryonic stem cells.

To directly address the contribution of chromatin modifications to gene expression and cellular identity, we have performed single-cell RNA sequencing of wild-type and various CRISPR-mediated knock-out mouse embryonic stem cell lines, generated in the same genetic background (Figure 1A). This set includes a line where all DNA methyltransferases were mutated (*Dnmt1,3a, 3b*-TKO) (Domcke et al., 2015), a cell line lacking H3K27me3 (*Eed-/-*) (Villasenor et al., 2020), and a cell line lacking the H3K36me3 methyltransferase SETD2 (Baubec et al., 2015). In addition, we generated a cell line lacking DNA methylation and H3K36me3 by mutating *Setd2* in the *Dnmt*-TKO background (here termed quadruple-KO, QKO), to test the combinatorial contribution of these marks on gene expression (Supplementary Figure S1A-D). All cell lines retained pluripotency and were able to proliferate without apparent morphological changes in serum + LIF conditions (Supplementary Figure S2A-E). We performed single-cell RNA sequencing on ~ 300 sorted cells for each line using SORT-seq (Hashimoto et al., 2012; Maurano et al., 2015) and obtained ~ 280 cells, with an average of 6000-7000 genes detected per cell after quality control and filtering (Supplementary Figure S3A and Material and Methods).

By visualising the transcriptional relationship of all cells using the uniform approximation and projection method (UMAP), we observed that removal of chromatin marks does not result in individual clusters for all individual mutant lines (Figure 1B), but rather forms a heterogenous cloud of cells, with *Eed-/-* cells as the only mutant line forming a separate cluster. Lack of DNA methylation results only in a minor separation from wild-type cells, while absence of H3K36me3 in the *Setd2-/-* cells does not result in any separation from the background line (both in *Setd2* single knock-out and QKO cells) (Figure 1B). This limited separation of epigenetic mutants is also observed using Principal Component Analysis (PCA, Supplementary Figure S3B) and reflects the limited changes in gene expression upon removal of H3K27me3, DNA methylation, or H3K36me3 (Supplementary Figure S2E, S3C and Table 1), in accordance with previous results obtained from bulk RNA-seq in mES cells (Boyer et al., 2006; Tsumura et al., 2006; Chamberlain et al., 2008;

Baubec et al., 2015). Nevertheless, we were able to identify known genes that change expression upon removal of DNA methylation or H3K27me3 such as upregulation of *Meg3* and *Dazl1* in TKO/ QKO mES cells or *Hoxd13* and *Lefty2* in *Eed-/-* ES cells (Supplementary Figure S3C).

Next, we wanted to quantify the consequences of chromatin modification removal on gene expression noise (cell-to-cell variability). We calculated the coefficient of variation (CV) for each gene in the individual cell lines and used pathway and gene set overdispersion analysis (PAGODA) (Fan et al., 2016) to adjust the variance for mean expression levels (Supplementary Figure S4A). We further collected genome-wide ChIP-seq datasets generated in bulk wild-type ES cells of identical background and analysed the association of these chromatin landmarks with the quantified adjusted variance of overlapping genes in wild-type ES cells. We observe the previously-described gene expression noise relationships with chromatin marks, transcriptional activity and CpG density at promoters and gene bodies - namely elevated noise at Polycomb-regulated genes and low variability at genes with CpG island promoters and H3K36me3 at gene bodies (Figure 1C and Supplementary Figure S4B-C) (Faure et al., 2017; Kar et al., 2017). Extending this analysis to DNA methylation, indicates a moderate increase in gene expression noise from genes with high promoter methylation. Next, we asked if the observed noise-chromatin associations are altered upon removal of the chromatin marks. Towards this we quantified the adjusted coefficient of variation in the cell lines lacking DNA methylation, H3K27me3 or H3K36me3, and compared this to wild-type cells. We did not observe any significant changes in cell-to-cell variation upon removal of these marks, indicating that there is no causal relationship between chromatin modifications and the observed transcriptional noise in pluripotent stem cells (Figure 1C and Supplementary Figure S4C).

Figure 1

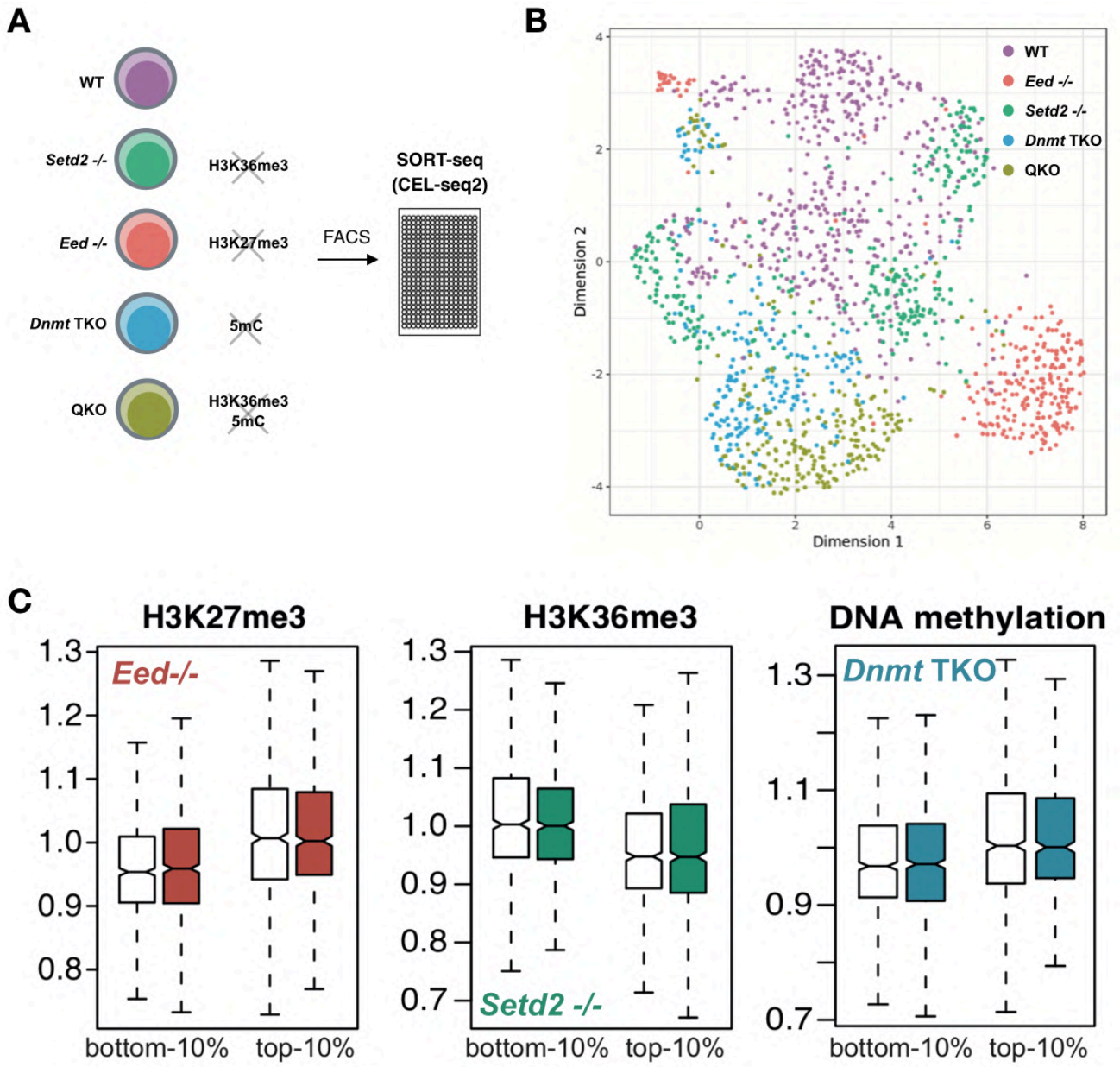


Figure 1 - Chromatin modifications do not influence gene expression noise in murine stem cells.

(A) Schematic overview of ES cell lines subjected to SORT-seq procedure. *Setd2-/-* cells lack all H3K36me3, *Eed-/-* lack H3K27me3, *Dnmt* TKO lack DNA methylation (5mC), and QKO (*Setd2-/-* in *Dnmt* TKO) cells lack both H3K36me3 and 5mC. FACS - fluorescence-activated cell sorting. **(B)** Visualisation of the transcriptional relationship between the different ES cell lines with the uniform manifold approximation and projection (UMAP) method for single-cell RNA-seq data. Each dot represents one cell. **(C)** Boxplots showing the gene expression noise relationship with different chromatin marks (regions with top/bottom 10 % occupancy) and changes upon loss of those. Shown is elevated noise at Polycomb-regulated genes (H3K27me3) and low variability at genes with CpG island promoters (DNA methylation) as well as H3K36me3 at bodies of protein-coding genes. See also Supplementary Figure S4.

3.2. Chromatin modifications influence exit from pluripotency and differentiation.

We next wanted to compare the contribution of these chromatin modifications to the establishment of new gene expression programs during exit from pluripotency and lineage commitment. Towards this, we first removed LIF from the medium to induce exit from pluripotency. Immediately after LIF withdrawal, chromatin-modification-deficient cells showed a significant decrease in cell survival, as described in previous studies (Jackson et al., 2004; Boyer et al., 2006; Sakaue et al., 2010; Zhang et al., 2014) (Figure 2A, Supplementary Figure S5A). Therein, *Eed-/-* cells were compromised stronger in comparison to the other mutant cell lines with a cell death of over 60 % (Supplementary Figure S5A). *Setd2-/-* cells showed increased cell death after the first passage, but prolonged passaging resulted in an increased survival rate from ~40 % to 80 % (Figure 2A).

Removal of LIF leads to uncontrolled differentiation of stem cells towards multiple lineages, therefore we wanted to compare the consequences of chromatin modification loss using a controlled differentiation system. Towards this, we differentiated the ES cells to neuronal progenitors and to mature glutamatergic neurons following an established *in vitro* differentiation protocol (Bibel et al., 2007) (Figure 2B-D, 2F, Supplementary Figure S5). As expected, cell lines lacking DNA methylation and H3K37me3 failed to reach the neuronal progenitor stage and did not give rise to post-mitotic neurons (Supplementary Figure S5B, S5D). The decrease in survival over time was accompanied by morphological changes of *Eed-/-* and TKO/QKO cellular aggregates such as hollow structures and frayed edges, indicating cellular detachment and increased cell death (Supplementary Figure S5C). In contrast, we observed that the *Setd2* knock-out cells did not show any differences compared to wild-type cells up to the neuronal progenitor stage (NPC), suggesting that exit from pluripotency and early lineage commitment is not affected in absence of H3K36me3 (Figure 2B, Supplementary Figure S5C). This was reproduced in two independent knock-out clones and one constitutive *Setd2* shRNA knock-down cell line (Figure 2B). However, further differentiation of *Setd2*-deficient progenitors to mature neurons was strongly impaired, with only 20 % of post-mitotic neurons surviving, indicating a potential role of SETD2/H3K36me3 in terminal differentiation or maintenance of neuronal identity (Figure 2C-D).

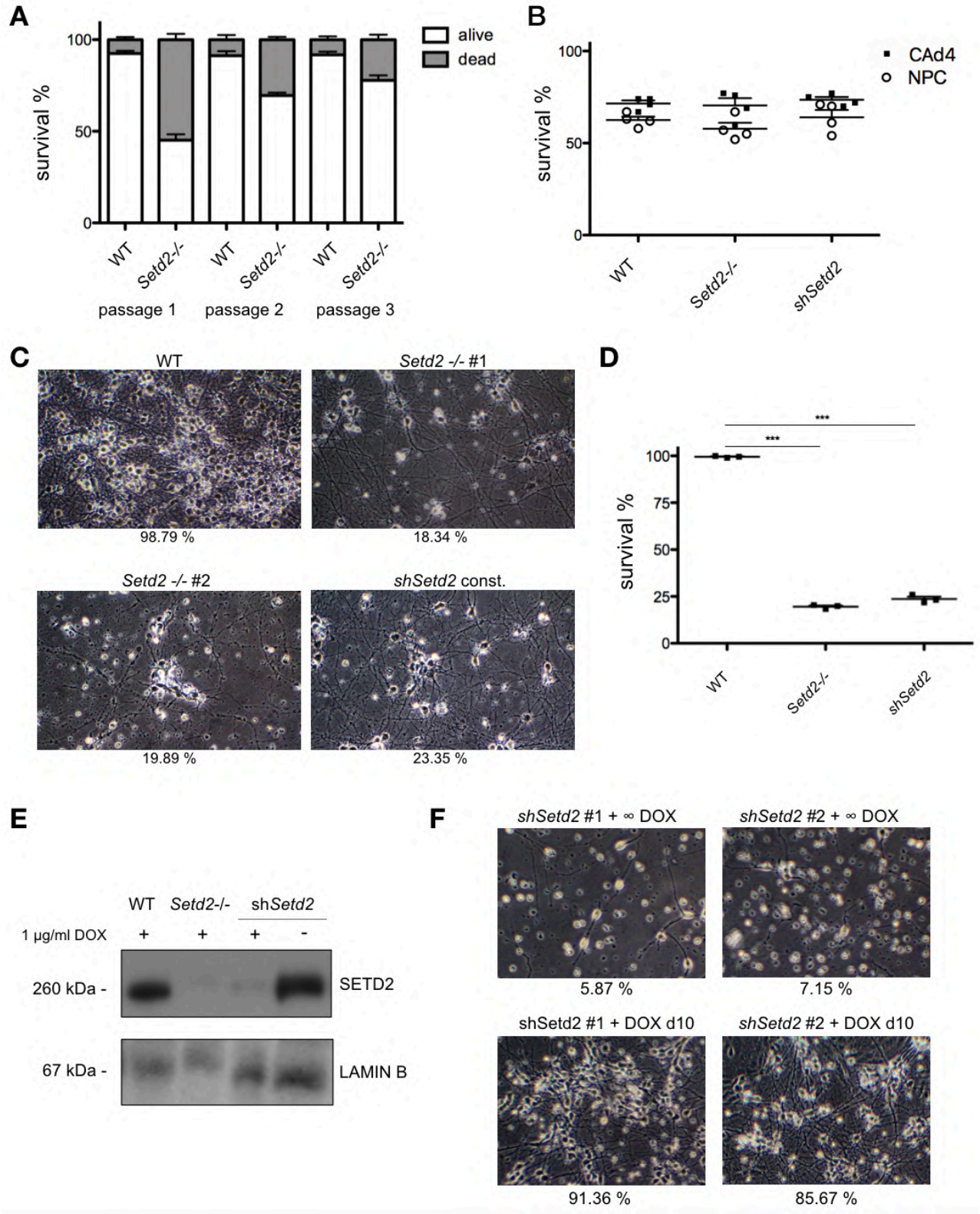
Following these observations, we wanted to explore the contribution of SETD2/H3K36me3 during the transition from progenitor to fully differentiated neurons. First, we asked if the reduced generation of neurons is due to impaired establishment or failure to maintain neuronal cell identity in absence of H3K36me3. To test this, we established a Tet-inducible *Setd2* shRNA knock-down

(sh*Setd2*) mES cell line and repeated the differentiation protocol while inducing shRNA expression at different timepoints (Supplementary Figure S6A). Loss of SETD2 and H3K36me3 in post-mitotic neurons was controlled by RT-qPCR, Western Blot and ChIP-seq (Figure 2E, Supplementary Figure S6B, S6D). Cells that were treated constantly or from early stages on (day d6, d8) with doxycycline showed a lower cell survival and low number of differentiated neurons, similar to the knock-out cells. In contrast, induction immediately after the early steps of neuronal differentiation (day d10) did not affect cell survival with 86-91 % surviving neurons (Figure 2F, Supplementary Figure S6C). This data indicates that SETD2 and H3K36me3 are necessary for lineage commitment, but dispensable once the neuronal identity has been established.

Figure 2 - Chromatin modifications influence exit from pluripotency and differentiation.

(A) Wild-type (WT) and *Setd2-/-* mES cells were cultured without leukemia-inhibitory factor (-LIF) over three passages. Shown are percentages of survival (life/dead stain) of three independent experiments. **(B)** Cell count assay using live-dead stain at stage of cellular aggregates (CA) day 4 and day 8 during embryoid body formation. Depicted are percentages of survival in WT, *Setd2-/-* and *Setd2* knockdown (sh*Setd2*) cells. **(C)** Microscopy images of *in vitro*-derived neurons of WT, *Setd2* -/- and constitutive *Setd2* knockdown cells at day d14. Shown are percentages of survival after dissociation (plated/attached) of three independent experiments; 100x magnification. **(D)** Survival rate (plated/attached) of terminal neurons of wild-type, *Setd2* -/- and constitutive *Setd2* knockdown cells at day d10; n=3. Statistical significance assessed by Student's t-test (p*** < 0.0001). **(E)** Representative microscopy images of *in vitro*-derived neurons of two Tet-inducible *Setd2* knockdown cell lines. Cells were treated with 1 µg/ml doxycycline (DOX) continuously or from day d10 on. Shown are percentages of survival after dissociation (plated/attached) of three independent experiments. **(F)** Immunoblot analysis for SETD2 levels in nuclear extracts of WT, *Setd2 -/-* and Tet-inducible sh*Setd2* knockdown terminal neurons at day d14 (+/- 1 µg/ml DOX). LAMIN B served as loading control.

Figure 2


A
survival %
100
50
0
alive
dead
WT
Setd2-/-
WT
Setd2-/-
WT
Setd2-/-
passage 1
passage 2
passage 3
B
survival %
100
50
0
CAd4
NPC
WT
Setd2-/-
shSetd2
C
WT
Setd2 -/- #1
98.79 %
18.34 %
Setd2 -/- #2
shSetd2 const.
19.89 %
23.35 %
D
survival %
***
***
100
75
50
25
0
WT
Setd2-/-
shSetd2
E
WT
Setd2-/-
shSetd2
1 µg/ml DOX
+
+
+
-
260 kDa -
SETD2
67 kDa -
LAMIN B
F
shSetd2 #1 + ∞ DOX
shSetd2 #2 + ∞ DOX
5.87 %
7.15 %
shSetd2 #1 + DOX d10
shSetd2 #2 + DOX d10
91.36 %
85.67 %

3.3. *Setd2-/-* compromises gene expression during terminal differentiation.

SETD2-mediated H3K36me3 is suggested to be involved in the regulation of transcriptional fidelity across the genome (Sen et al., 2015; Faure et al., 2017; Meers et al., 2017). Thus, we assessed the transcriptome of *Setd2-/-* neuroprogenitors, the cell state before cell death. PolyA-RNA sequencing analysis showed minor changes in gene expression in *Setd2*-deficient NPCs with ~300 differentially expressed genes (Figure 3A). Gene ontology analysis of down-regulated genes showed high enrichment for neurodevelopmental-related processes, suggesting a deregulation of neuronal transcriptional programs already at this stage (Figure 3B). Given these mild, but detectable changes in the overall transcriptome and remaining 20 %-survival rate upon terminal differentiation, we wanted to investigate if absence of H3K36me3 results in increased cellular and gene expression heterogeneity at the progenitor stage. We therefore performed single-cell RNA-sequencing at the NPC stage. Visualising the transcriptional relationship of all cells using the UMAP method showed that removal of H3K36me3 results in a separation of *Setd2-/-* from wild-type NPCs (Figure 3C), in agreement with transcriptional differences observed from bulk RNA-seq.

Next, we wanted to quantify the consequences of H3K36me3 removal on gene expression noise (cell-to-cell variability). We calculated the coefficient of variation (CV) for each gene in the individual cell lines and used PAGODA (Fan et al., 2016) to adjust the variance for mean expression levels (Figure 3D). Comparison of CV values with genome-wide ChIP-seq datasets generated in bulk wild-type and *Setd2-/-* NPCs showed again similar trends as observed in ES cells, such as higher variability at H3K27me3-regulated promoters (Figure 3E, Supplementary Figure S7A-B). However, these associations were still present in cells lacking H3K36me3, supporting that H3K36me3 does not contribute to cell-to-cell variation in gene expression of NPCs.

Given the changes in transcription for neurodevelopmental genes in *Setd2-/-* NPCs in our bulk RNA-seq data (Figure 3A-B) and the observed separation in the single-cell data (Figure 3C), we further assessed the expression of neuronal transcription factors in the single-cell RNA-seq data and by RT-qPCR in NPCs (Figure 3F, Supplementary Figure S7C and D). We did not only observe the impaired upregulation of major neuronal transcription drivers (e.g. *Pax6, Pax3, Olig2*, *Nestin*), but also the significant upregulation of wrong differentiation drivers (e.g. *Sox17*, *Cdx2*) as well as a insufficient downregulation of pluripotency marker genes (*Nanog, Pouf5F1*), the latter indicating an incomplete or a delay in differentiation. This data suggests a role of SETD2-mediated H3K36me3 in fine-tuning the correct establishment of gene expression programs.

Figure 3

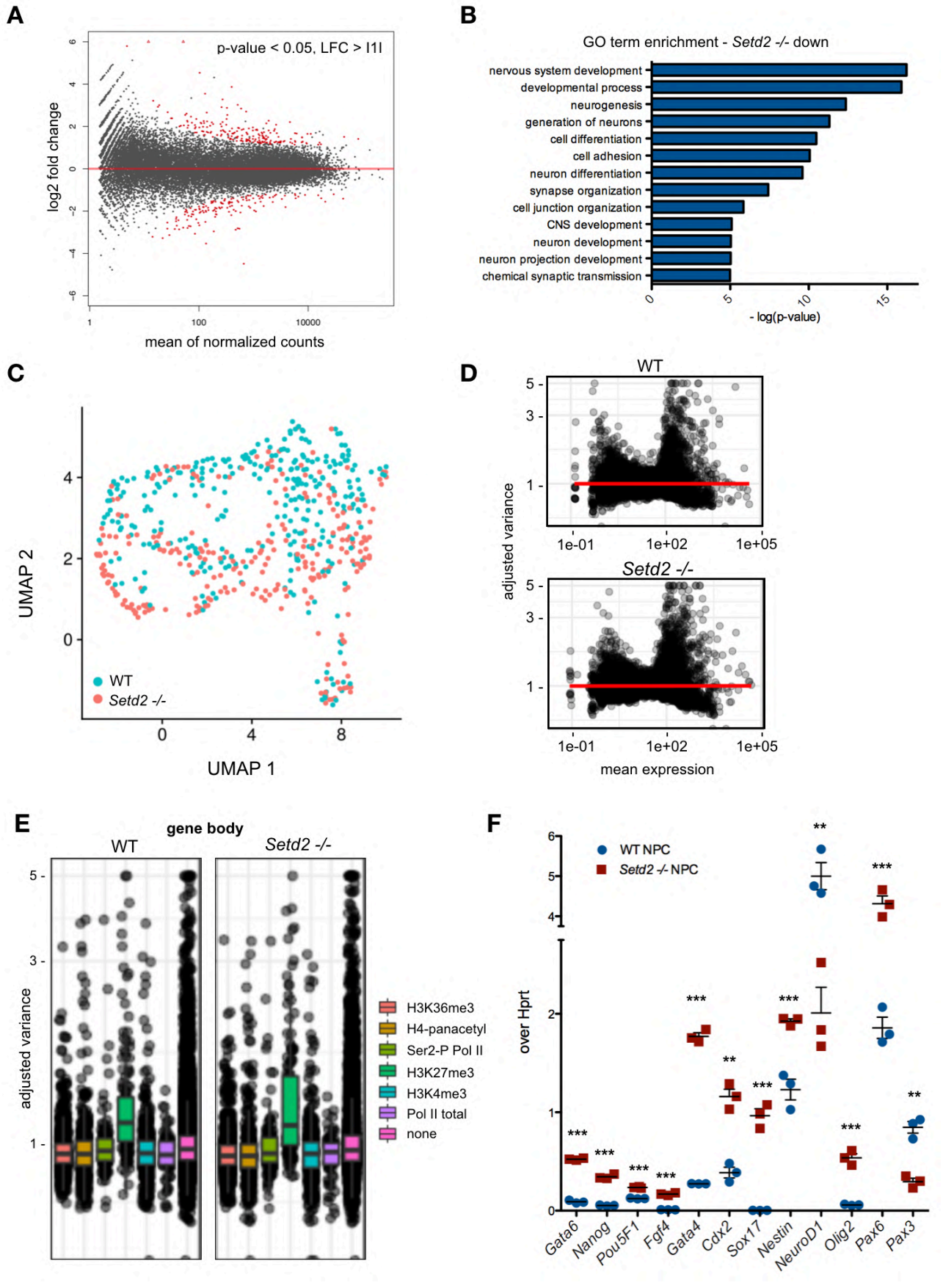

A
p-value < 0.05, LFC > I1I
log2 fold change
mean of normalized counts
B
GO term enrichment - Setd2 -/- down
nervous system development
developmental process
neurogenesis
generation of neurons
cell differentiation
cell adhesion
neuron differentiation
synapse organization
cell junction organization
CNS development
neuron development
neuron projection development
chemical synaptic transmission
- log(p-value)
C
UMAP 2
UMAP 1
WT
Setd2 -/-
D
WT
Setd2 -/-
adjusted variance
mean expression
E
gene body
WT
Setd2 -/-
adjusted variance
H3K36me3
H4-panacetyl
Ser2-P Pol II
H3K27me3
H3K4me3
Pol II total
none
F
WT NPC
Setd2 -/- NPC
over Hprt
Gata6
Nanog
Pou5F1
Fgf4
Gata4
Cdx2
Sox17
Nestin
NeuroD1
Olig2
Pax6
Pax3

Figure 3 - Chromatin modifications influence exit from pluripotency and differentiation.

(A) DESeq2 MA plot of bulk PolyA-RNA-seq results depicting differences in gene expression between wild-type (WT) and *Setd2-/-* NPCs (~350 DEGs, p-value < 0.05, LFC > I1I). **(B)** Gene ontology (PantherGO) enrichment analysis for down-regulated genes in *Setd2*-deficient NPCs. Shown are the most enriched GO terms with a -log(p-value) > 5. **(C)** Uniform approximation and projection (UMAP) method for single-cell RNA-seq data of WT and *Setd2-/-* neuronal progenitors. There is a mild cluster formation between the two cell lines. **(D)** Calculation of the coefficient of variation and its adjustment for mean expression levels for each gene in individual WT and *Setd2-/-* neuronal progenitor cells using the pathway and gene set overdispersion analysis (PAGODA). **(E)** Genome-wide signals at gene bodies (starting +2kb from TSS) of ChIP-seq datasets generated in bulk neuronal progenitor cells were associated with the quantified adjusted variance to analyse gene expression noise relationships with chromatin marks in WT and *Setd2-/-* NPCs. **(F)** RT-qPCR of various lineage marker analysis in WT and *Setd2-*/- NPCs (n=3). Statistical significance assessed by Student's t-test (p** < 0.01, p*** < 0.0001). *Hprt* - housekeeping gene, *Oct4 (Pou5f1)*, *Gata6*, *Nanog* - embryonic stem cell markers; *Nestin*, *NeuroD1*, *Pax3*, *Pax6*, *Olig2* - neuronal marker; *Fgf4* - growth factor, *Gata4* - extra-embryonic endoderm marker, *Cdx2* - trophoblast marker, *Sox17* - parietal endoderm marker.

3.4. SETD2/H3K36me3 ensure correct gene expression during differentiation.

To further address the impact of *Setd2-/-* in ES cell-derived neuronal progenitor cells, we measured the expression of the neuronal surface markers CD24/CD56 by FACS. We observed a significant decrease in the expression of these markers during the last stage of embryoid body formation, prior to the dissociation of NPCs. More than twice as many CD24-/CD56- double-negative NPCs were detected in the *Setd2-/-* population (~21 %) as in wild-type cells (~8 %) (Figure 4A-B), indicating a deficit in developmental progression for *Setd2-/-* cells. Prolonged cultivation of embryoid bodies in retinoic acid for four more days (CAd12) showed a small, but positive effect on the survival of *Setd2*-deficient cells (Supplementary Figure S8A). However, this increase in survival was also observed for the wild-type cells (Supplementary Figure S8B). Further, surface marker and RT-qPCR analysis at day 12 of cellular aggregation confirmed that *Setd2-/-* cells showed only a minor improvement of marker gene or surface marker (CD24-/CD56- double-negative: ~5 % in wild-type, ~22 % in *Setd2-/-* at CAd12) expression in comparison to WT or *Setd2-/-* cells at day 8 of cellular aggregation (Figure 3F, 4B, Supplementary Figure S7D, S8C-D). Thus, prolonged cultivation of progenitors minimally improves differentiation, but does not completely revert gene expression changes caused by the lack of SETD2/H3K36me3.

In order to assess if this deficit in developmental progression is dependent on failure to establish the correct gene expression program, we generated *Setd2-/-* ES cell lines with Tet-inducible expression of the two bHLH transcription factors that promote neurogenesis: *Neurogenin 1* and *Neurogenin 2* (i*Ngn*) (Material and Methods, Supplementary Figure S8E). Induced expression of these factors have been shown to enable rapid neurogenesis by inducing genetic programs involved in the transition from stem cells to pre-natal neurons (Busskamp et al., 2014). We argued that their expression in *Setd2-/-* cells could be sufficient to rescue incorrect establishment of gene expression programs in absence of H3K36me3 and to induce neurodifferentiation. Induction of i*Ngn* indeed increased the survival rate by ~30 %, leading to a partial rescue of the cell death observed in absence of SETD2/H3K36me3 (Figure 4C-D). In addition, RT-qPCR analysis for selected lineage marker genes confirmed this partial rescue on a molecular level, where induction of *Ngn1/2* in *Setd2-/-* cells showed a change in expression of lineage markers similar to wild-type cells (Figure 4E). This correction of the transcriptional program in comparison to wild-type NPCs (+DOX) is not complete; however, it is sufficient for an overall improvement in cell survival. This data suggests that SETD2-mediated H3K36me3 ensures accurate transcriptional programming upon cellular differentiation and its loss leads to failure to establish correct gene expression during neuronal differentiation. Nevertheless, this deficit can be rescued by enforcing the correct transcriptional programs through induced expression of neuronal transcription factors.

Figure 4

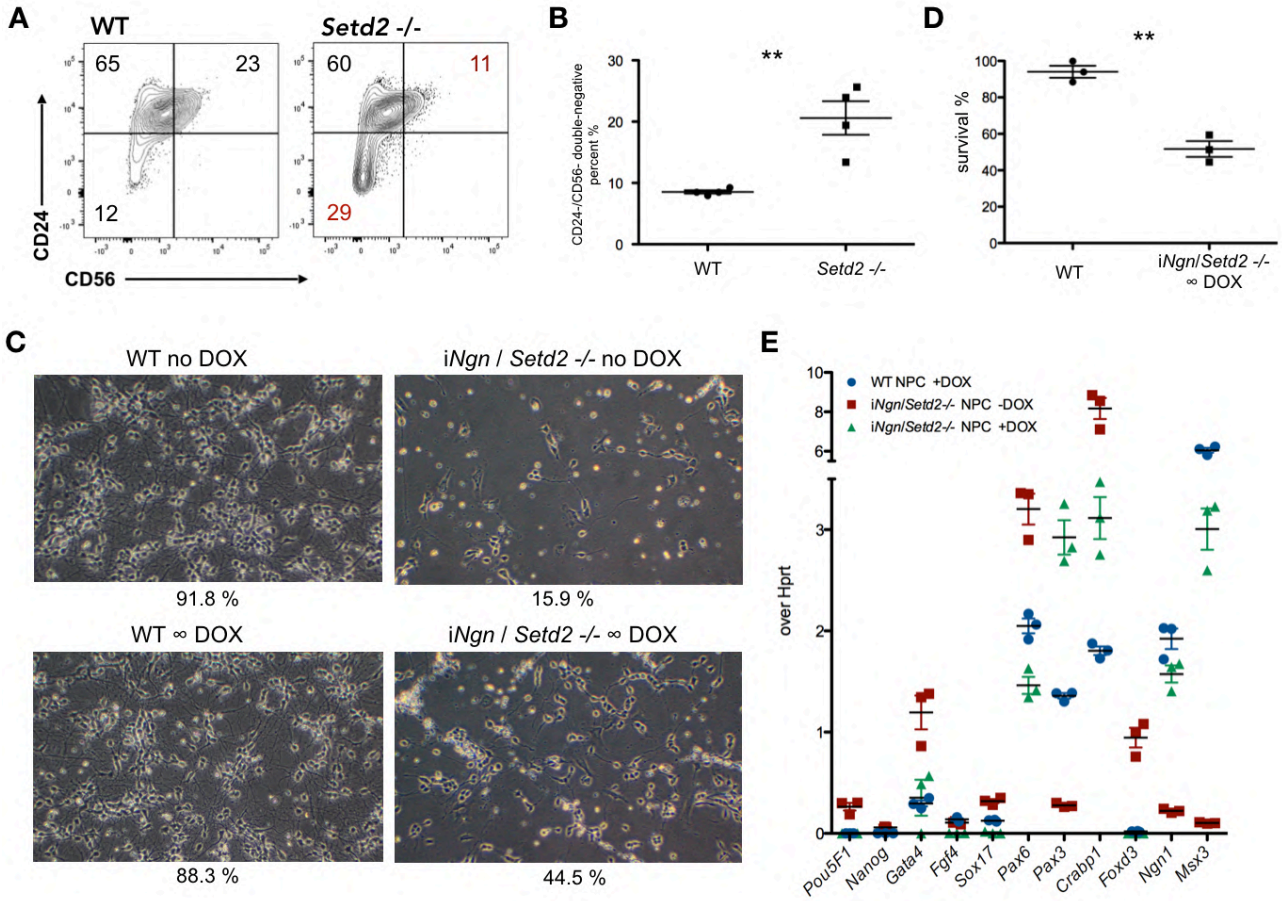


Figure 4 - *Setd2-/-* cell-derived NPCs exhibit an inaccuracy in lineage programming which is partially rescued by introduction of a major neurotranscription factor.

(A) - **(B)** Flow cytometry measurements for the surface markers CD24 and CD56 in wild-type (WT) and *Setd2-/-* ES cell-derived neuronal progenitor cells after embryoid body dissociation. **(A)** Exemplary flow cytometry dot plots. **(B)** Plot indicates the percentage of double-negative (CD24-/CD56-) NPCs (four independent measurements per sample are shown). Statistical significance assessed by Student's t-test (p** < 0.01). **(C)** Microscopy images of *in vitro*-derived neurons shows successful differentiation of neuronal cells derived from WT control ES cells and partial rescue of cell death in *Setd2*-deficient cells by inducible overexpression of *Neurogenin2-Neurogenin1* (i*Ngn*). Similar results were obtained from two independent replicates. Cells were either mock-treated or continuously treated with 1 µg/ml doxycycline (DOX) during *in vitro* differentiation. **(D)** Shown are percentages of survival after dissociation (plated/ attached) of three independent experiments described in (C). n=3. Statistical significance assessed by Student's t-test (p** < 0.01, p*** < 0.0001). **(E)** RT-qPCR of various lineage marker analysis in WT and *Setd2-/-* NPCs, the latter expressing *Neurogenin1* and *Neurogenin2* upon 1 µg/ml doxycycline treatment (n=3). *Hprt* - housekeeping gene, *Oct4 (Pou5f1)*, *Nanog* - embryonic stem cell markers; *Pax3*, *Pax6*, *Ngn1*, *Crabp1*, *Msx3*, *Foxd3* - neuronal marker; *Fgf4* - growth factor, *Sox17* - parietal endoderm marker.

In summary, we have identified a crucial role for SETD2-mediated H3K36me3 in the accurate establishment of gene expression during neuronal development. Given the readout of this histone modification and the role of SETD2 as a potential tumor suppressor as described in chapter 1.3., we wanted to further decipher the potential mechanism as to how this enzyme and its mark contribute to the establishment of gene expression profiles needed for terminal differentiation.

3.5. DNA Damage Response in Absence of SETD2/H3K36me3.

H3K36me3 has been shown to play a role in aiding DNA damage response and DNA repair pathway choice in actively transcribed regions (Pfister et al., 2014; Huang et al., 2018). Loss of SETD2-mediated H3K36me3 in cancer cells has been further linked to increased DNA damage and genomic instability (Skucha et al., 2018). Thus, we wanted to assess the potential role of elevated DNA damage levels in the low survival during differentiation of cells lacking SETD2/ H3K36me3. In collaboration with Fede Teloni (Altmeyer lab, UZH), we first investigated phosphorylation levels of the DNA damage marker γH2A.X in *Setd2-/-* mES cells using quantitative, image-based cytometry (QIBC) (Toledo et al., 2013). We found that even after challenging the cells with DSB-inducing agent Camptothecin, no differential DNA damage response between WT and knock-out ES cells could be detected (Figure 5A-B). We further investigated phosphorylation γH2A.X by immunoblotting using a γH2A.X-specific antibody in NPCs, the stage before cell death establishes. Again, we were not able to detect changes in the DNA damage response in *Setd2-/-* cells (Figure 5C), suggesting that the loss of H3K36me3 does not increase DNA damage in mouse ES cells and during their differentiation to neuronal progenitors.

Figure 5

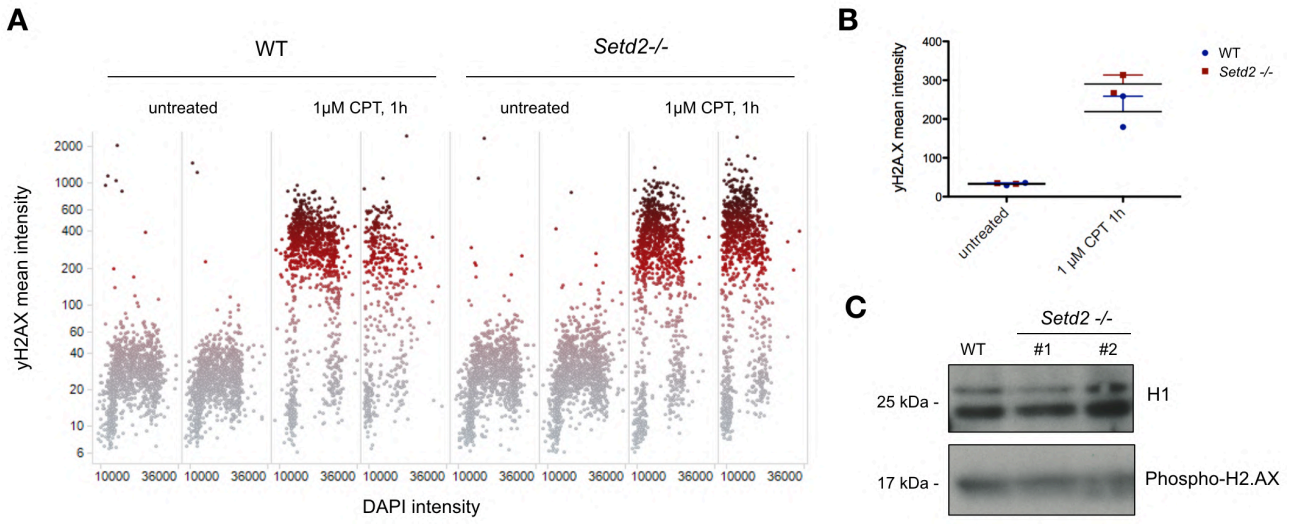


Figure 5 - H3K36me3 loss does not impair DNA Damage Response.

(A)-(B) Untreated or treated (1µM Camptothecin CPT for 1h) WT and *Setd2-/-* mES cells have been fixed, permeabilised, and stained for γH2AX level in a duplicate fashion. Automated fluorescence microscopy and follow-up normalisation of the Cy5 signal (γH2AX) to the DAPI signal (cell nuclei) of 1000 cells shows no significant differences in DNA damage response. **(A)** Mean Cy5 intensity of each cell in one sample. **(B)** Quantification of (A) showing average mean intensity of all cells in replicates, n=2. **(C)** Immunoblot analysis of serine 129 phosphorylation of H2A.X in untreated WT and *Setd2-/-* neuronal progenitors (NPCs). H1 serves as loading control.

3.6. Histone H3.3 Variant Incorporation at H3K36me3-high Sites.

H3.3, a histone variant that is structurally very close to the canonical histone H3, has been associated with genomic regions that undergo elevated chromatin-turnover, such as sites of active transcription (Buschbeck and Hake, 2017). Its role in histone replacement at active genes is highly conserved and has been proposed to participate in the transmission of active chromatin states (Orsi et al., 2009). H3.3 is co-transcriptionally incorporated in gene bodies, therefore tri-methylation on lysine 36 is predominantly found on histone H3.3 (H3.3K36me3), and specific reader proteins have been identified that recognise H3.3K36me3 (Wen et al., 2014).

Since H3.3 becomes enriched in maturing post-mitotic cells, it is particularly important for neuronal plasticity (Maze et al., 2015). Thus, the increased incorporation of H3.3 into actively transcribed gene bodies during neuronal differentiation could be influenced in absence of H3K36me3, potentially leading to a differential downstream readout and altered gene expression (Wen et al., 2014). To test this hypothesis, we studied histone H3.3 localisation by introducing an exogenous biotin-tagged H3.3 protein to a heterologous site (Figure 6A) in wild-type and *Setd2-/-* ES cells. Chromatin immunoprecipitation experiments followed by sequencing (ChIP-seq) in mES cells and NPCs did not show drastic differences in H3.3 levels at H3K36me3-rich gene bodies (Figure 6B), suggesting that H3.3 localisation is not influenced by H3K36me3 loss.

Figure 6

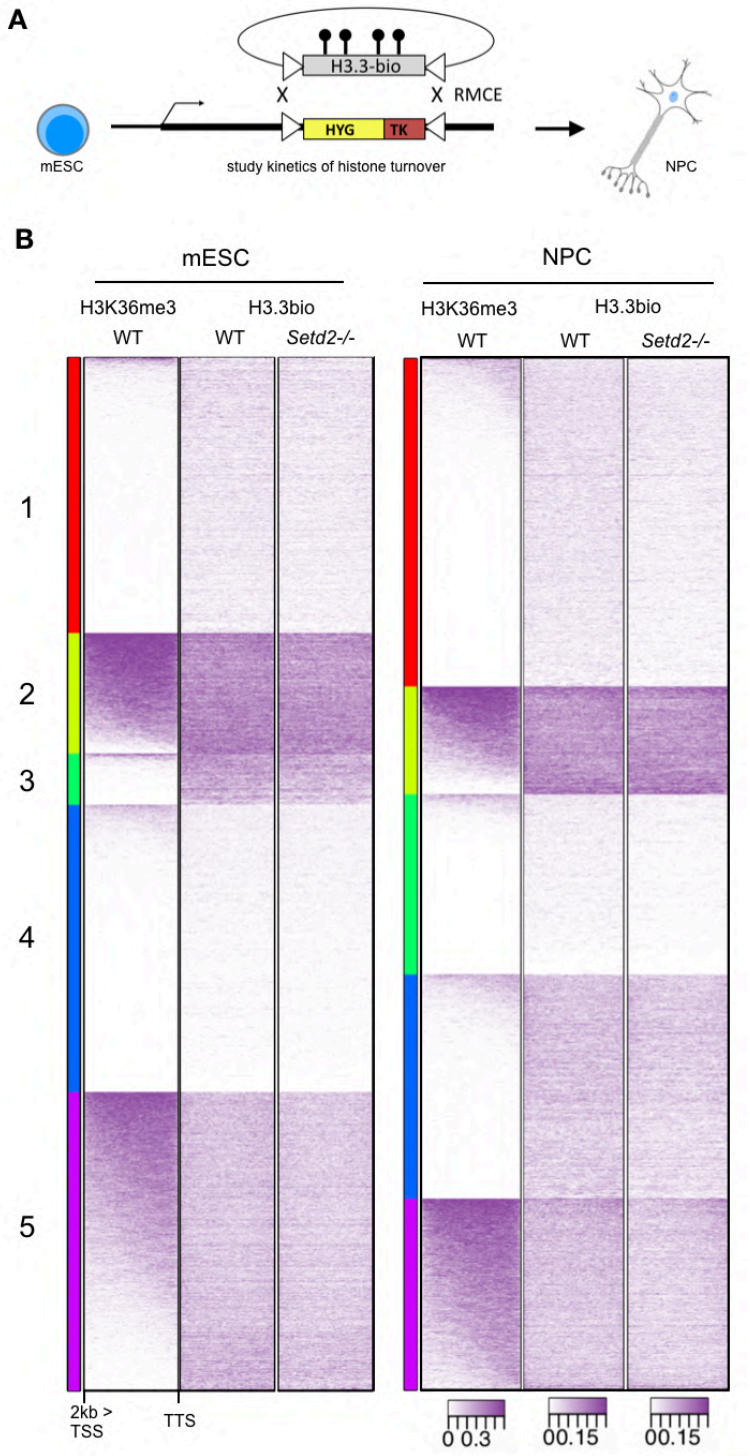


Figure 6 - No differential H3.3 levels at formerly H3K36me3-high gene bodies.

(A) Schematic overview of the generation of H3.3-bio overexpressing ES cells. H3.3 gene was inserted with a biotin-acceptor peptide sequence into BirA ligase and RMCE-competent ES cells. Positive clones were further differentiated to neuronal precursors (NPC). HYG - hygromycin, TK - thymidine kinase. **(B)** Clustered heatmaps showing H3.3 bio-ChIP-seq signals in gene bodies, starting 2 kilobases from transcription start (TSS) until transcription termination (TTS) site) of WT vs. *Setd2-/-* mES cells and NPCs. Signals were ranked according to H3K36me3 ChIP-seq signals in WT cells.

3.7. Influence of H3K36me3 loss on the Chromatin Landscape.

Binding analysis by ChIP-seq for H3K27me3, RNA Pol II as well as Ser2-P Pol II and H4-panacetylation have been performed in late neuronal progenitors to elucidate as to what extent these are influenced by lack of H3K36me3, and if these changes are associated with the altered transcriptional output observed in *Setd2-/-* NPC (Figure 7). We focused our analysis of H3K36me3-rich gene bodies, starting 2 kilobases from the TSS, as changes should be expected there upon depletion of SETD2/H3K36me3.

We did not observe large rearrangements of these chromatin modifications upon loss of H3K36me3 could in *Setd2-/-* late neuronal precursors. However, at formerly H3K36me3-enriched sites, a slight decrease in Pol II signal and increase in H4-panacetyl signal can be observed (Figure 7A-B). This is in line with the observed role of H3K36me3 in yeast, where it promotes recruitment of the histone deacetylase complex Rpd3S to transcribed gene bodies (Carrozza et al., 2005). Since histone acetylation is known to be an active marker for open chromatin and is directed by H3K36me3 in gene bodies (Zhang et al., 2015b), we performed accessibility analysis using ATAC-seq in WT and *Setd2-/-* NPCs (Figure 7C). We did not observe differentially accessible sites between WT and mutant NPCs, indicating that there is a slight increase of H4 acetylation in gene bodies, but it does not affect chromatin accessibility or Pol II occupancy.

Figure 7

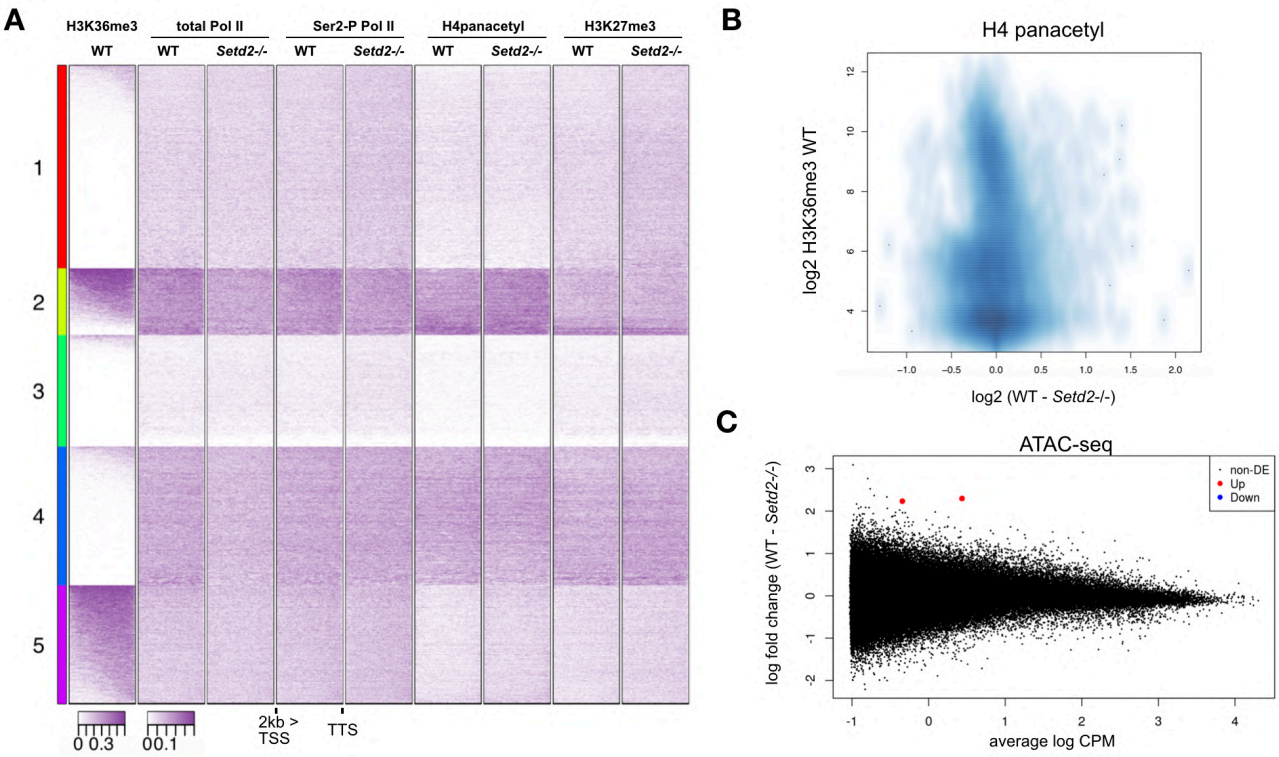


Figure 7 - Chromatin Environment in WT vs. *Setd2-/-* neuronal precursors.

(A) Clustered heatmaps showing antibody ChIP-seq signals for H3K36me3, total and Ser2-P Pol II, H4 panacetyl in gene bodies (2kb from TSS until transcription termination site) of WT vs. *Setd2-/-* NPCs. Signals were ranked according to H3K36me3 ChIP-seq signals in WT NP cells. **(B)** Smooth scatter plot of delta genic H4panacetyl ChIP-seq signals (WT-*Setd2-/-*) over genic H3K36me3 ChIP-seq-signal in WT NPCs. All values are log2-transformed. There is a slight increase in acetylation at gene bodies that loose H3K36me3 in *Setd2-/-* cells. **(C)** Accessibility changes were assessed by ATAC-seq in WT vs. *Setd2-/-* NPCs. Shown is a MA plot for called ATAC-seq peaks in WT vs. *Setd2-/-* NPCs in dependence on average counts per million (CPM). All values are log2 transformed. Red dots show significantly different accessible sites (logFC > 0.5, FDR < 0.05).

3.8. Changes in Protein-Chromatin Interactions in Presence and Absence of SETD2/H3K36me3.

H3K36me3 could have a role as a bookmark to mark chromatin of transcribed genes and regulate recruitment of important factors to these sites (see chapter 1.3.2.). Thus, we wanted to investigate potential changes in the protein constitution on chromatin in absence of H3K36me3 and its enzyme SETD2 (Figure 8). First, we wanted to analyse potential changes of global chromatin interactions in absence/presence of H3K36me3 (Figure 8A). Towards this, we performed global chromatin mass spectrometry experiments in collaboration with the Stengel lab (University of Konstanz). Stable isotope labelling of WT and *Setd2*-/- mES cells with amino acids in cell culture (SILAC) followed by mass spectrometry (MS) analysis of cross-linked, chromatin-bound fractions (ChEP) (Kustatscher et al., 2014) were performed to obtain a quantitative and unbiased detection of protein-chromatin interaction changes in absence of SETD2 (Supplementary Figure S9A). The results revealed additional factors influenced by the lack of SETD2/H3K36me3, but less than expected from their potential role in RNA synthesis and genome function. As a control hit, we could rely on KDM5B a H3K4me2/3 demethylase that has been shown to associate to H3K36me3 sites in mES cells (Xie et al., 2011) and is depleted in the *Setd2-/-* background. In general, Pol II subunits were not differentially enriched between WT and *Setd2-/-* cells, suggesting that Pol II occupancy on chromatin itself is not altered upon H3K36me3 loss in ES cells. This was further validated by immunoblotting for Pol II protein levels in chromatin-bound fractions of WT and *Setd2-/-* mES cells (Supplementary Figure S9B).

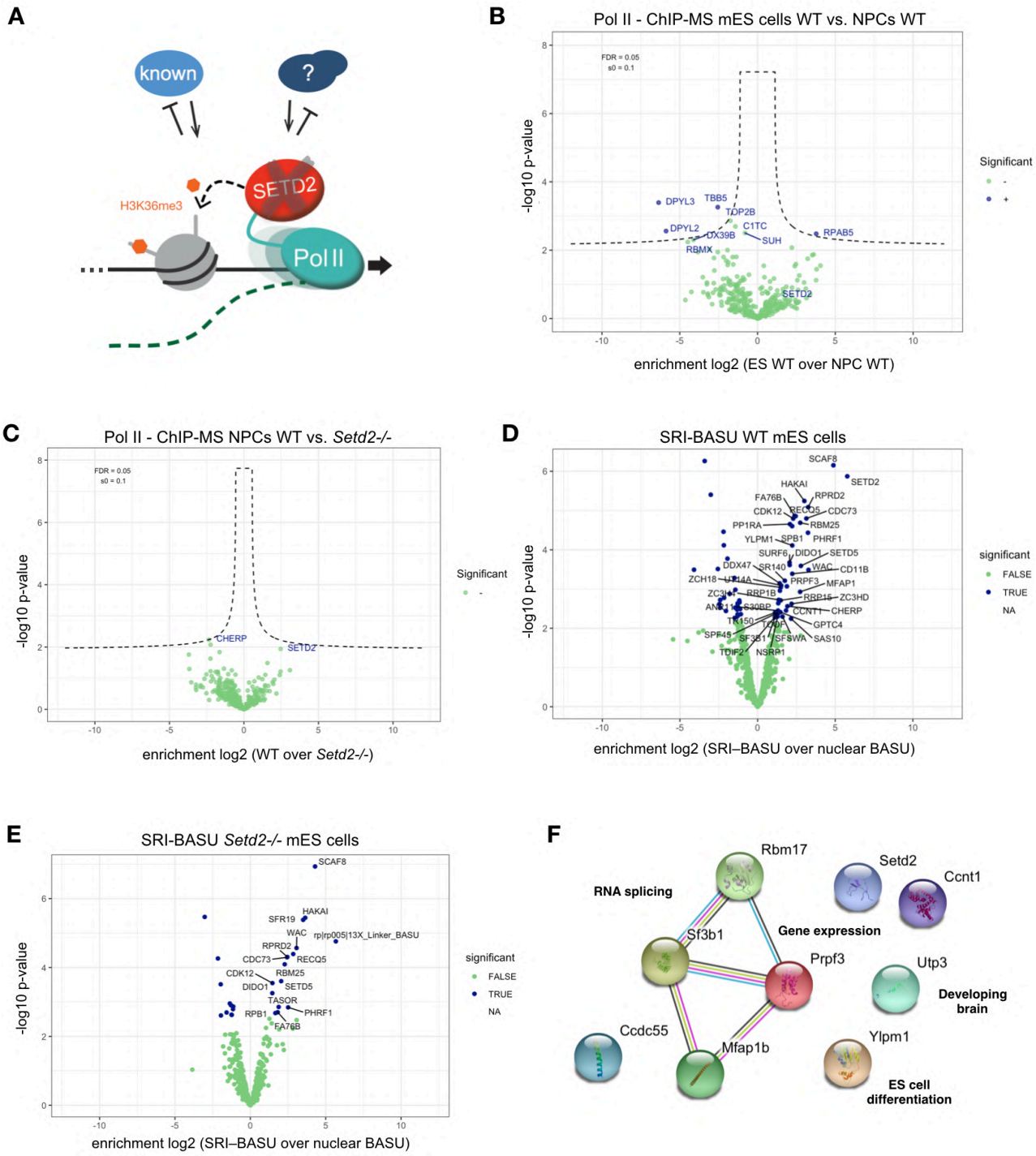
We next wanted to focus on the interactome of actively transcribed gene bodies, rather than the entire genome. Thus, we performed ChIP experiments against serine-2-phosphorylated (Ser2-P) Pol II followed by MS detection of associated proteins. Between wild-type ES and NP cells, the Pol II interactome only slightly changes by enrichment of neuronal-specific factors such as the Dihydropyrimidinase Like proteins 2 and 3 (DPYL2/3) in NPCs (Figure 8B). Changes between wild-type and *Setd2-/-* NPCs could not be observed (Figure 8C), indicating that the interactome at actively transcribed genes is not altered upon H3K36me3 loss. However, this antibody-based approach holds limitations. It might not be representative for the protein-chromatin interactions around H3K36me3, as passaging by the elongating Pol II disrupts present chromatin states and potential interactions. Further, since this approach involves cross-linking with formaldehyde, capturing short-lived or dynamic changes in the chromatin-protein interactome might be limited (Solomon and Varshavsky, 1985).

Due to these limitations, we utilised the recently established tool ChromID (Villasenor et al., 2020) and the SRI domain of SETD2 which binds to the Ser2-P CTD of the elongating Pol II. A promiscuous BirA* ligase from *Bacillus subtilis* (BASU) was added to the latter and successfully introduced into wild-type and *Setd2-/-* mES cell background for proximity labelling experiments using mass spectrometry. Biotinylation of proteins was sufficient leading to a count of about 4000 peptides in the different samples (Supplementary Figure S9C). Enrichment analysis of SRI-BASU over a non-directed BASU control in wild-type ES cells showed an increase in the overall captured protein interactome (Figure 8D) in comparison to the previous Ser2-P-Pol II-MS run (Figure 8B), calling 39 significantly enriched proteins (Supplementary Figure S9D). Further, in the *Setd2-/-* background 18 of those significantly enriched proteins were lost besides SETD2 itself and 5 new interactions were gained, indicating a potential role for SETD2/H3K36me3 in facilitating interactions with the elongating RNA polymerase (Figure 8E). Next, we subjected a selection of these differentially enriched proteins, which are lost upon SETD2 depletion to GO term enrichment analysis (Supplementary Figure S9E) and to the STRING database (Figure 8F) (Szklarczyk et al., 2019). We could find an enrichment of proteins being involved in 3’ RNA processing, particularly mRNA splicing. These results suggest that the observed transcriptome changes could be due to altered interactions between RNA processing factors and chromatin.

Figure 8 - Protein-Chromatin Interactome in Presence and Absence of SETD2/H3K36me3.

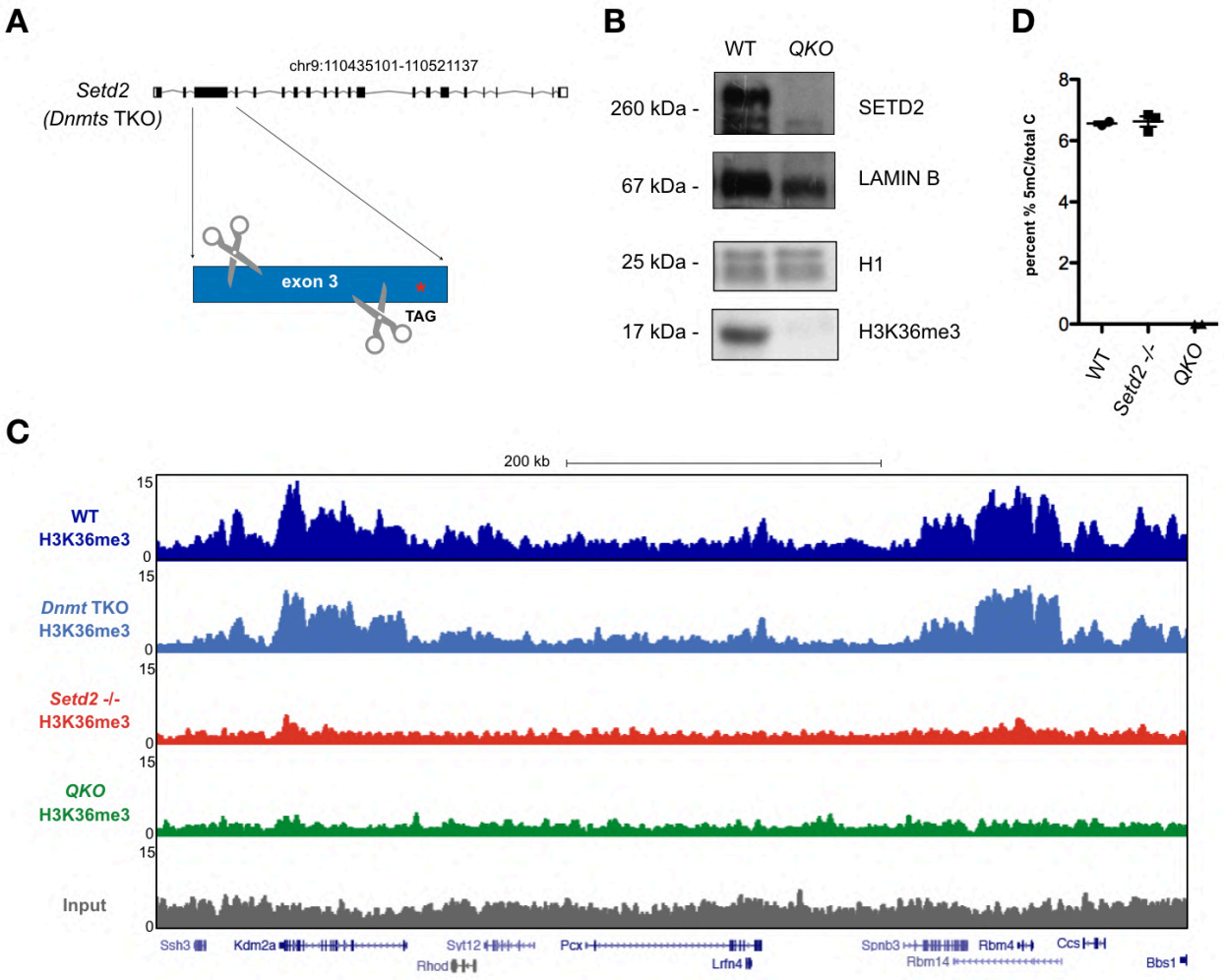
(A) Schematic overview of potential identification of known as well as unknown protein interactions facilitated by SETD2, H3K36me3 and the elongating Pol II. **(B)-(C)** Volcano plot showing Ser2-P Pol II ChIP-MS results obtained in (B) WT ES and NP cells or (C) WT and *Setd2-/-* NP cells. There are no statistically enriched proteins (FDR = 0.01, s0 = 0.1, log2 FC > 0, *n* = 4). **(D)-(E)** Volcano plots indicating ChromID results from SETD2-SRI domain engineered chromatin reader-BASU fusion in comparison to a reader-free BASU control in (D) WT or (E) *Setd2-/-* ES cells. Statistically enriched proteins are indicated in blue (FDR = 0.01, s0 = 0.1, log2 FC > 0, n = 4). **(F)** Analysis of selected proteins lost in *Setd2-/-* SRI-BASU ES cells using the STRING database.

Figure 8


A
known
?
SETD2
H3K36me3
Pol II
B
Pol II - ChIP-MS mES cells WT vs. NPCs WT
FDR = 0.05
s0 = 0.1
-log10 p-value
enrichment log2 (ES WT over NPC WT)
Significant
DPYL3
TBB5
TOP2B
DPYL2
DX39B
C1TC
SUH
RPAB5
RBMX
SETD2
C
Pol II - ChIP-MS NPCs WT vs. Setd2-/-
CHERP
SETD2
enrichment log2 (WT over Setd2-/-)
D
SRI-BASU WT mES cells
enrichment log2 (SRI–BASU over nuclear BASU)
significant
FALSE
TRUE
NA
E
SRI-BASU Setd2-/- mES cells
F
RNA splicing
Rbm17
Sf3b1
Gene expression
Prpf3
Setd2
Ccnt1
Utp3
Developing brain
Ccdc55
Mfap1b
Ylpm1
ES cell differentiation

4. Results - Supplementary Information

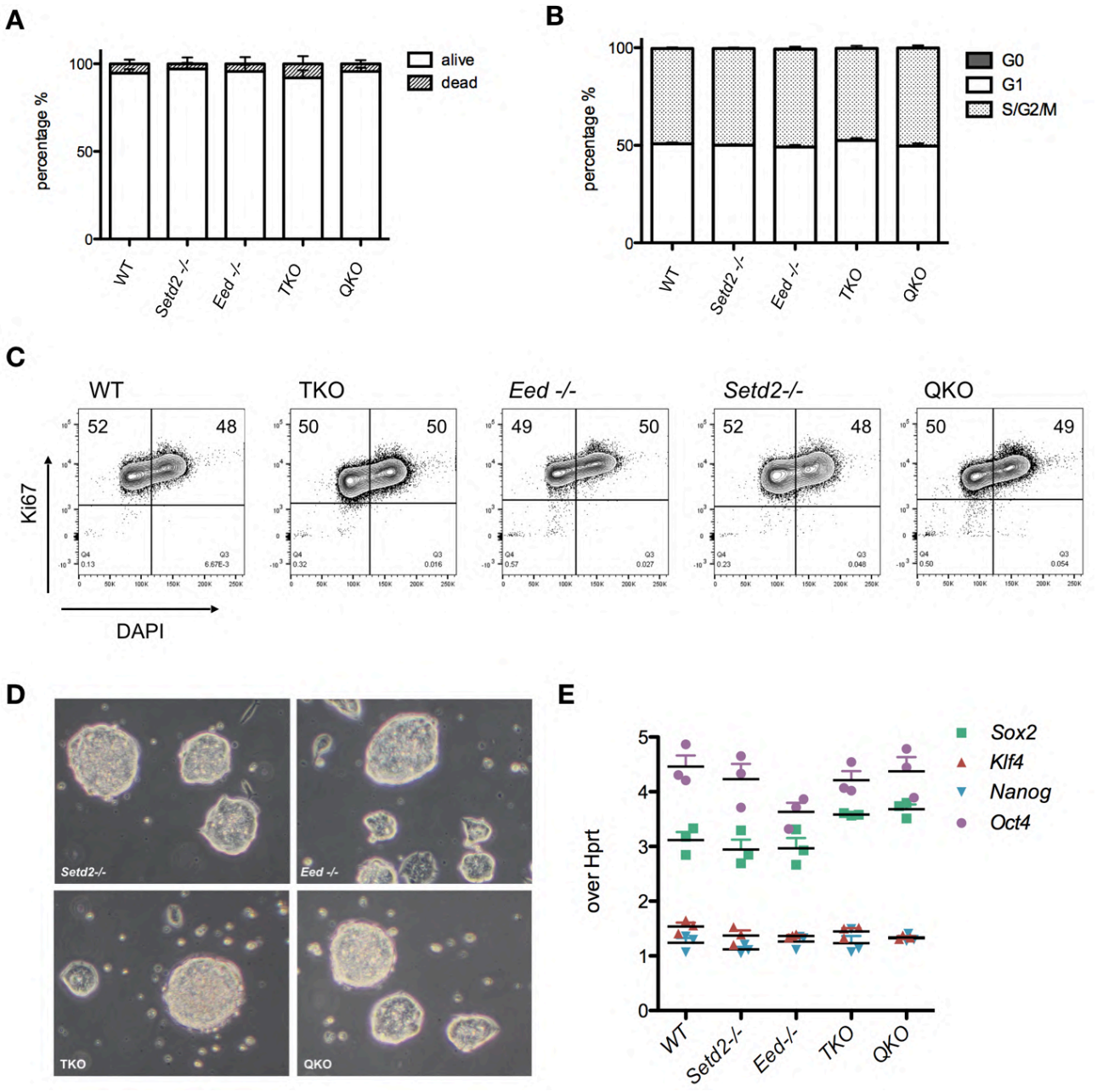
Supplementary Figure S1



Supplementary Figure S1 - Generation of *Setd2-/-* ES cells in a Dnmt TKO background.

(A) CRISPR/Cas9 knock-out strategy for *Setd2* in *Dmnt* TKO background (QKO). Two guide RNAs target exon 3, resulting in an out-of-frame deletion and a downstream premature STOP codon. **(B)** Immunoblot analysis for SETD2 and H3K36me3 levels in nuclear (top) and histone (bottom) extracts of wild-type (WT) and QKO mES cells. LAMIN B and H1 serve as loading controls. **(C)** Representative genome browser view of a chromosome 19 locus exemplifying differences in signal for H3K36me3 between wild-type, *Dnmt* TKO, *Setd2 -/-*, and QKO mES cells. Shown are read counts per 100 bp for ChIP-seq and input samples. **(D)** HPLC-MS measurement of methylcytosine to cytosine indicates absence of methylation in the QKO. Error bars denote standard deviation from three independent replicate measurements.

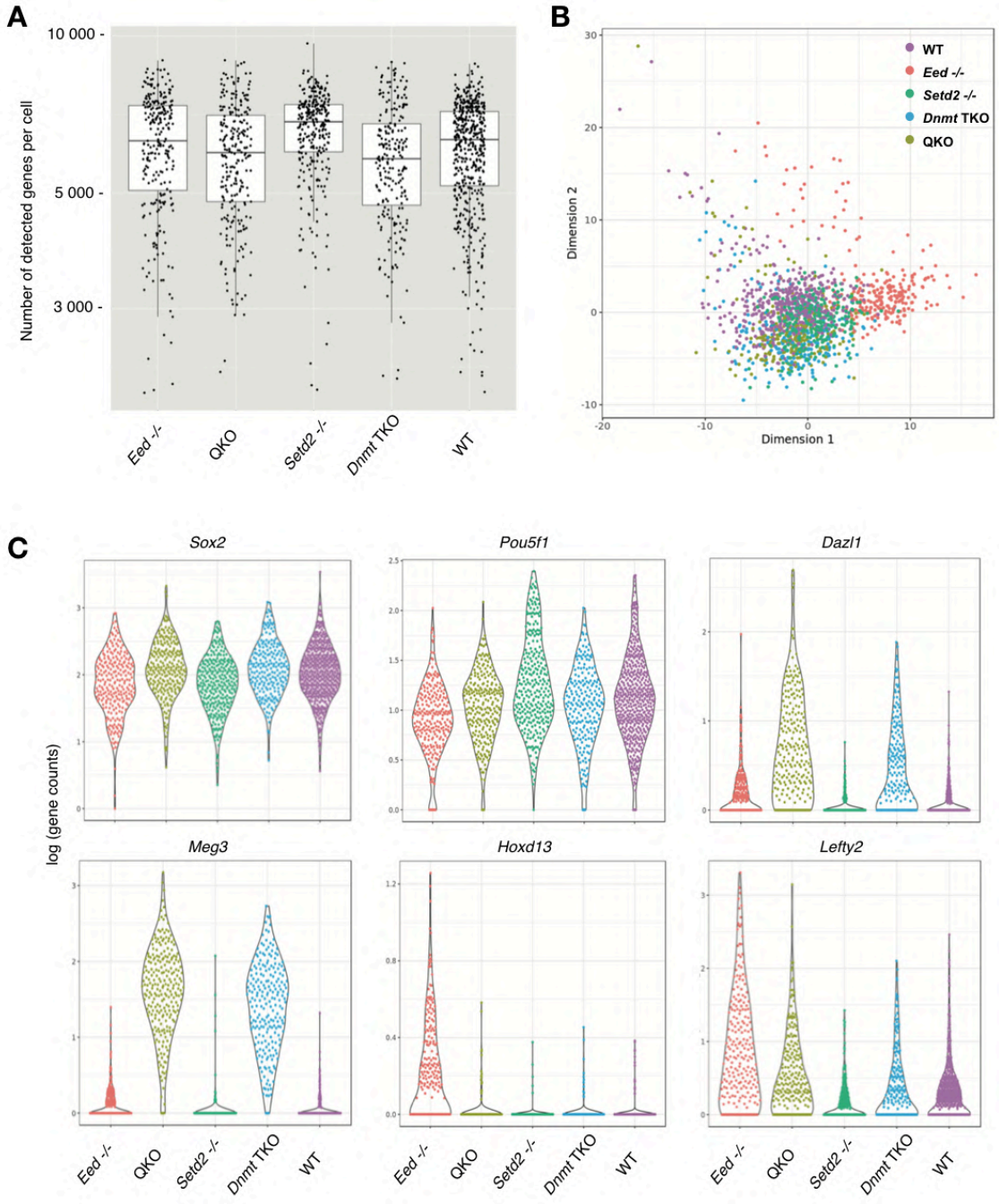
Supplementary Figure S2



Supplementary Figure S2 - Loss of chromatin marks does not impair self-renewal capacity and proliferation of mES cells.

(A) Wild-type (WT), *Setd2 -/-*, *Eed -/-*, *Dnmt* TKO and QKO mES cells were cultured in media containing leukemia-inhibitory factor (+LIF) over three passages. **(B)** Ki-67 and DAPI cell cycle analysis by FACS. Shown are percentages of the different phases of WT, *Setd2 -/-*, *Eed -/-*, *Dnmt* TKO and QKO mES cells, n = 3. **(C)** Exemplary FACS analysis for Ki-67 and DAPI levels in WT, *Setd2 -/-*, *Eed -/-*, *Dnmt* TKO and QKO mES cells. Highlighted are percentages of cells in Ki67+ (G1 phase) and Ki67+/DAPI+ (S/G2/M phase) clusters. **(D)** Microscopy images of *Setd2 -/-*, *Eed -/-*, *Dnmt* TKO and QKO mES cells in feeder-free culture. 100x magnification. **(E)** Real-time qPCR detection of pluripotency marker genes *Nanog*, *Pou5f1* (*Oct4*), *Klf4*, and *Sox2 in* WT, *Setd2 -/-*, *Eed -/-*, *Dnmt* TKO and QKO mES cells. *Hprt* serves as an internal control.

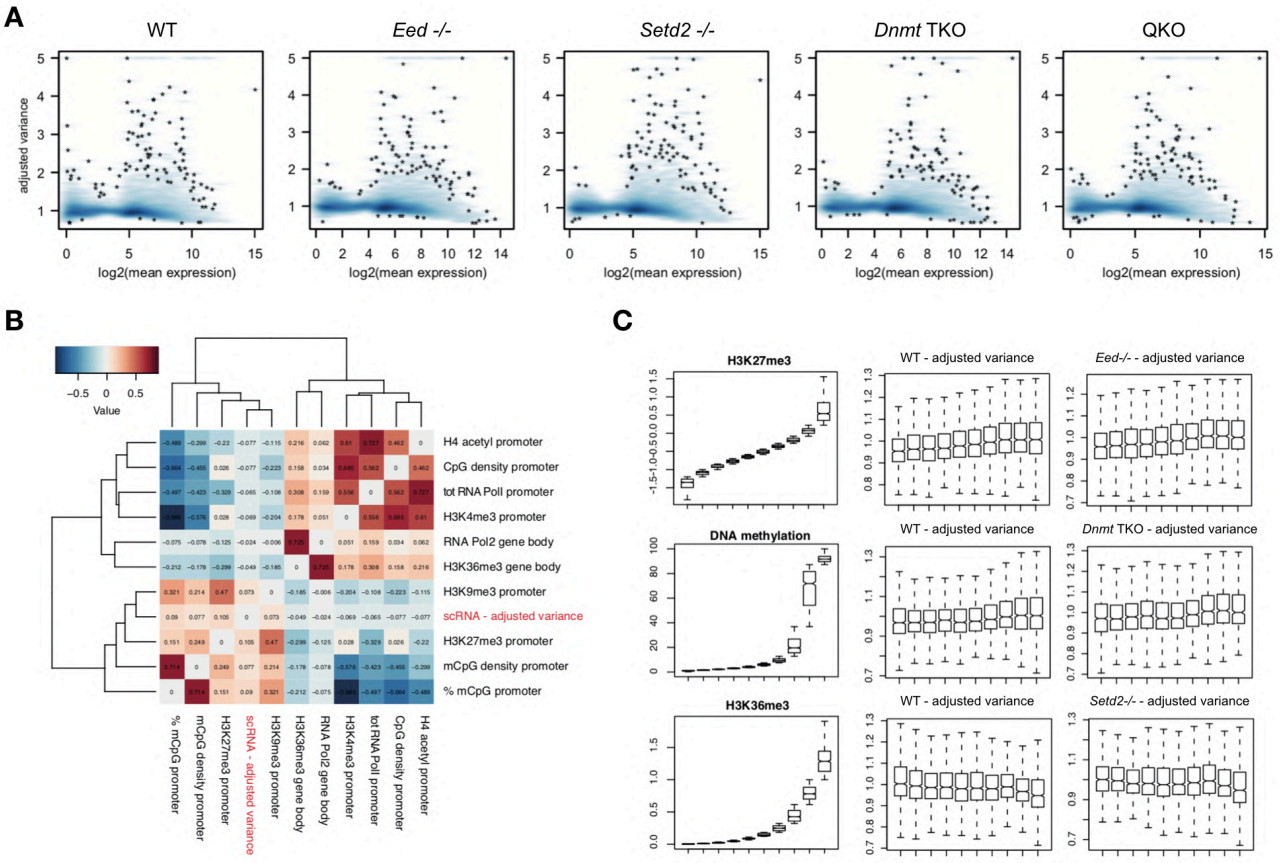
Supplementary Figure S3



Supplementary Figure S3 - Differential Gene Expression in Wild-type, *Setd2-/-*, *Eed-/-*, *Dnmt* TKO and QKO mES cells.

(A) Boxplot showing the number of detected genes per single cell in Wild-type (WT), *Setd2-/-*, *Eed-/-*, *Dnmt* TKO and QKO mES cells. Counts extracted after removal of cell cycle effects, of cells with high mitochondrial gene content, and of genes with no counts. **(B)** PCA plot showing cluster of *Eed-/-* mES cells in comparison to random spreading of WT, *Setd2-/-*, *Eed-/-*, *Dnmt* TKO and QKO mES cells. **(C)** Violin plots showing the log-transformed gene counts for selected genes in single WT, *Setd2-/-*, *Eed-/-*, *Dnmt* TKO and QKO mES cells.

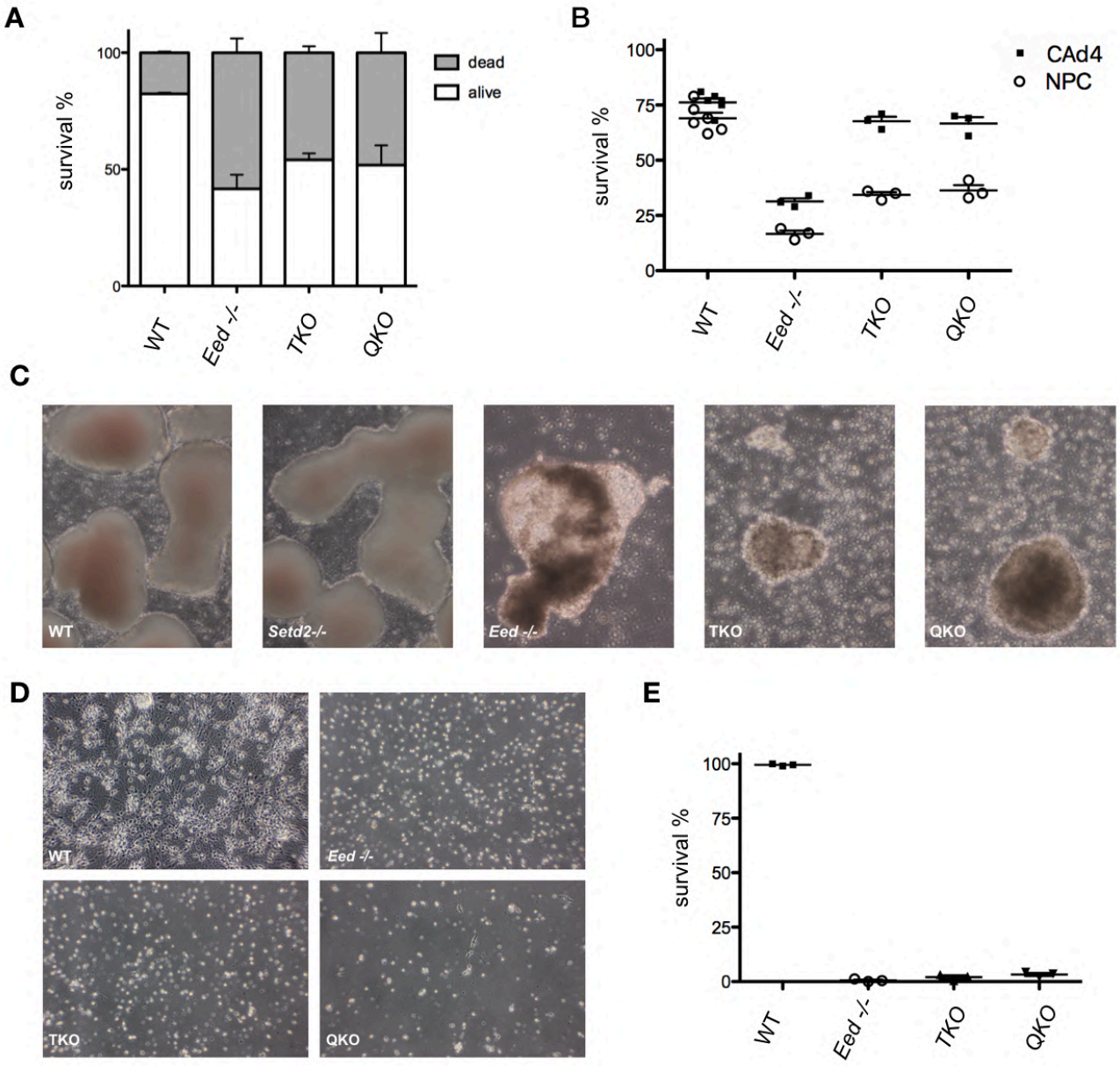
Supplementary Figure S4



Supplementary Figure S4 - Differential Gene Expression in Wild-type, *Setd2-/-*, *Eed-/-*, *Dnmt* TKO and QKO mES cells.

(A) PAGODA framework to adjust the variance for mean expression levels after calculation of the coefficient of variation for each gene in the individual cell lines. **(B)**-**(C)** Gene expression variability analysis integration with chromatin modifications data using normalised ChIP-seq data summarised by gene for signals at the TSS (+/- 2 kb from the transcriptional start site) and in gene bodies (+2 kb from transcriptional start site). **(B)** Heatmap presenting the correlation scores between gene expression variability and various ChIP-seq signals. **(C)** DNA methylation and H3K27me3 levels at promoters or H3K36me3 levels in gene bodies were quantilised into deciles (left). Normalised gene expression variability in wild-type and corresponding mutant mES cell lines across the obtained deciles is shown on the right.

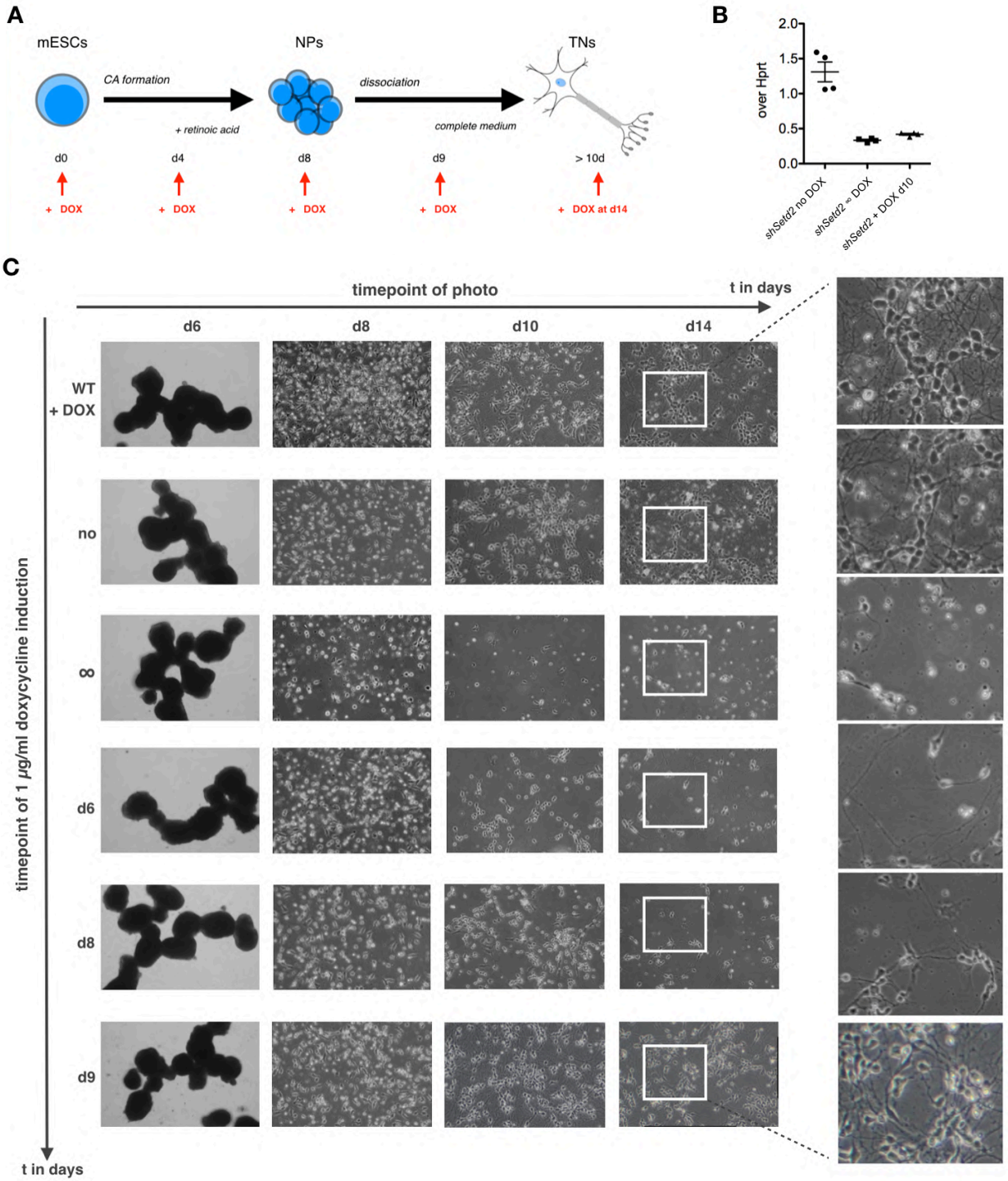
Supplementary Figure S5



Supplementary Figure S5 - *Eed-/-*, *Dmnt* TKO and QKO lead to loss of differentiation potential.

(A) Wild-type (WT), Eed-/-, Dmnt TKO and QKO mES cells were cultured without leukemia-inhibitory factor (-LIF). Shown are percentages of survival based on life-dead stain after the first passage in three independent experiments. **(B)** Cell count assay using live-dead stain at CA day 4 and NPC stage. Depicted are percentages for survival in WT, *Eed-/-, Dmnt* TKO and QKO cells. **(C)** Exemplary microscopy images of in vitro-derived cellular aggregates (CA) at day 4 for wild-type, *Eed-/-, Dmnt* TKO and QKO cells. 100x magnification. **(D)** Exemplary microscopy images of in vitro-derived, terminal neurons of wild-type, *Eed-/-, Dmnt* TKO and QKO cells at day 14. Shown are percentages of survival (plated/attached); 100x magnification. **(E)** Cell count assay using live-dead stain at terminal neuron stage d14. Depicted are percentages for survival in WT, *Eed-/-, Dmnt* TKO and QKO cells.

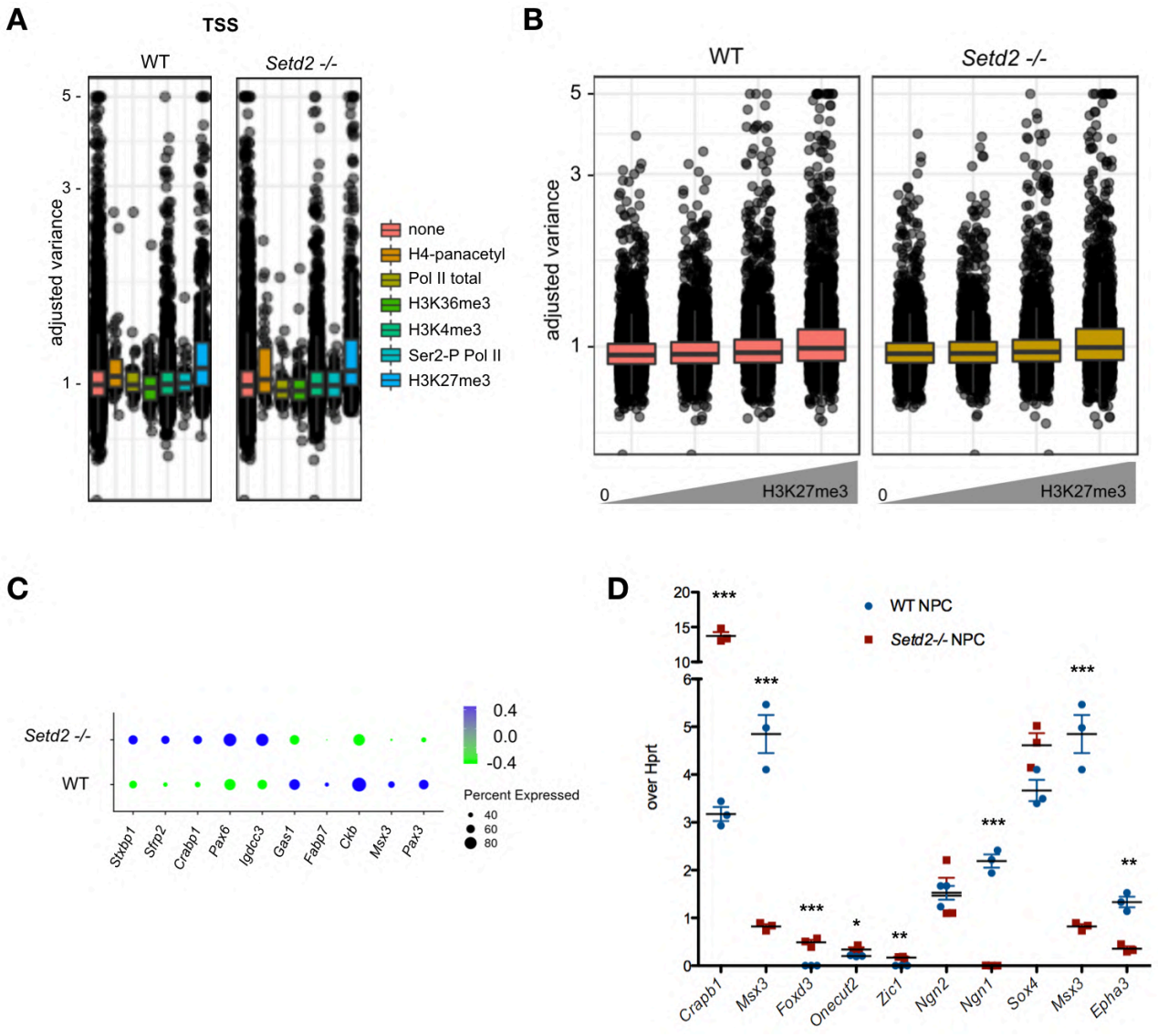
Supplementary Figure S6



Supplementary Figure S6 - Neurodifferentiation upon induction of a *Setd2* knockdown.

(A) Schematics depicting the neuronal differentiation protocol (Bibel et al. 2007). Red arrows indicate start of 1 µg/ml doxycycline (DOX) induction over time. **(B)** Real-time qPCR of *Setd2* in Tet-inducible *Setd2* knockdown cells, differentiated to terminal neurons (day d14). Cells were treated with 1 µg/ml doxycycline, always or from day d10 on. *Hprt* was used as an internal control. **(C)** Time-resolved microscopy images of *in vitro* differentiation of tet-inducible *Setd2* knockdown cell lines, starting from cellular aggregates (CA) day 6 over neuronal progenitors at day 8 to terminal neurons at day 10 and 14. Treatment with 1 µg/ml doxycycline (DOX) was initiated at day 0, 6, 8, and 9; 100x magnification. **(D)** Representative genome browser view of a chromosome 7 locus exemplifying differences in ChIP-Rx signal for H3K36me3 between Wild-type (WT), *Setd2-/-*, and Tet-inducible *shSetd2 cells* at terminal neuron stage day d16. Shown are read counts per 100 bp for ChIP-seq samples. shRNA was induced at day d10 with 1 µg/ml doxycycline (DOX).

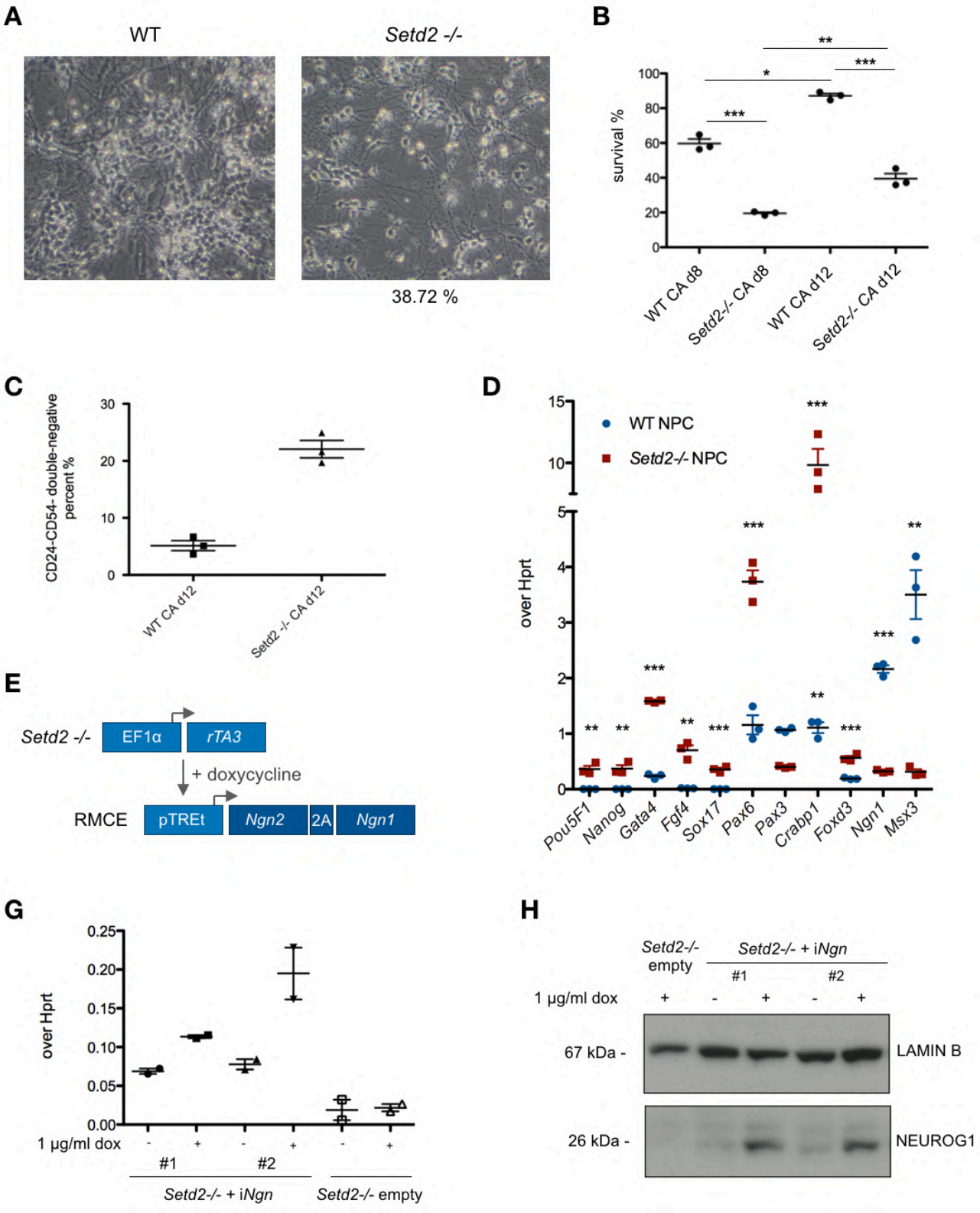
Supplementary Figure S7



Supplementary Figure S7 - Single-cell RNA-seq analysis of wild-type vs. *Setd2*-deficient neuronal progenitors.

(A) Genome-wide signals at the transcription start site (TSS, +/- 2kb) of ChIP-seq datasets generated in bulk neuronal progenitor cells were associated with the quantified adjusted variance to analyse gene expression noise relationships with chromatin marks in wild-type (WT) and *Setd2-/-* NPCs. **(B)** H3K27me3 ChIP-seq-centered analysis of adjusted variance in wild-type (WT) and *Setd2-/-* NPCs. H3K27me3-levels at TSS (+/- 2kb) were quantilised from lowest to highest, showing a mild dependency of the adjusted gene expression variance on H3K27me3 signal in NPCs. **(C)** Unsupervised detection of genes that separate WT and *Setd2-/-* NPCs based on single-cell RNA-seq. Shown are log-fold change ranges of the expression for each gene between the two cell lines. **(D)** Validation of differential expression of genes identified in single-cell RNA-seq data sets of WT and *Setd2-/-* NPCs by RT-qPCR analysis (n=3). *Hprt* served as internal control. Statistical significance assessed by Student's t-test (p** < 0.01, p*** < 0.0001).

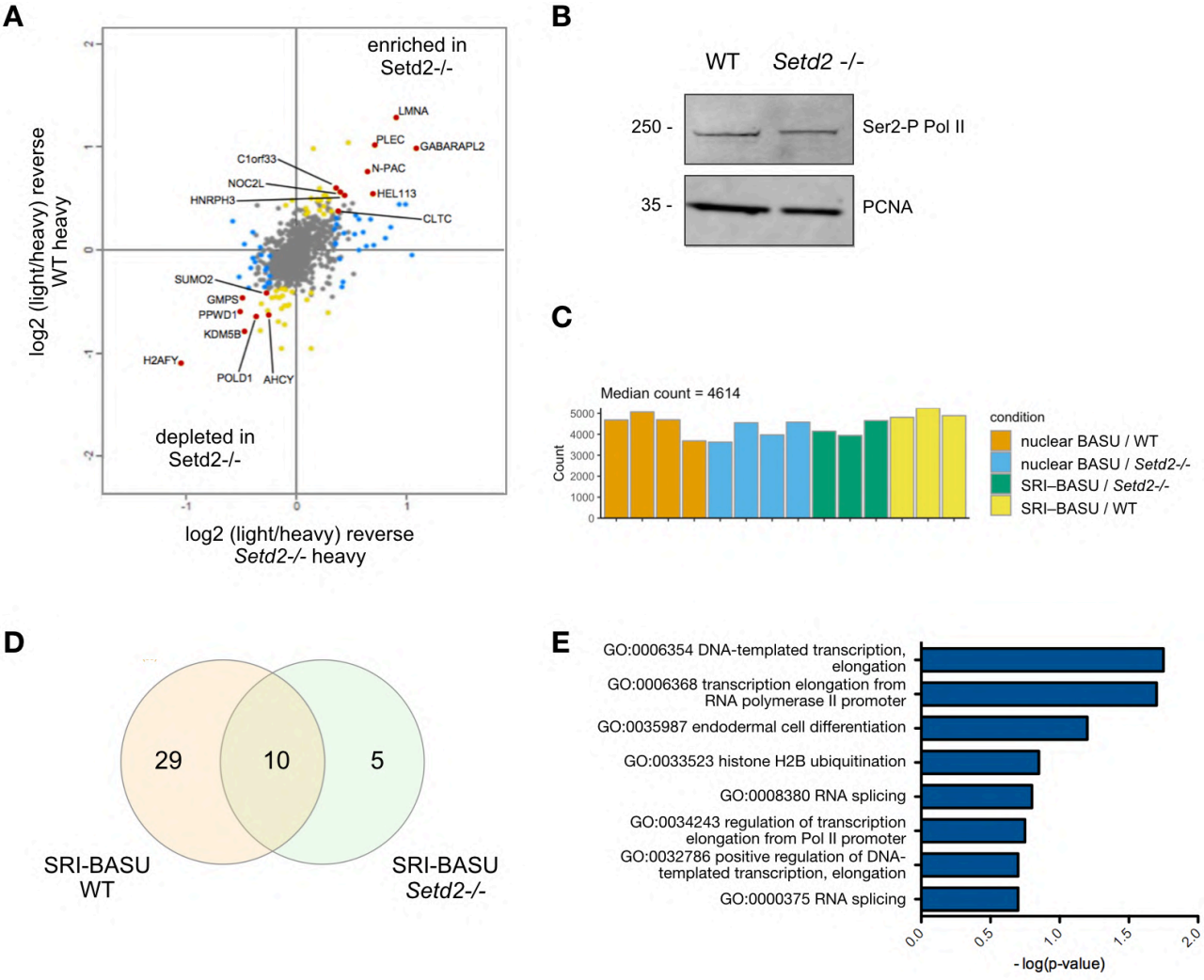
Supplementary Figures S8



Supplementary Figure S8 - Effects of prolonged cultivation in the late progenitor stage and overexpression of neuronal transcription factors on *Setd2*-deficient NPCs.

(**A)** Microscopy images of *in vitro*-derived neurons shows successful differentiation of neuronal cells derived from wild-type (WT) control ES cells and a partial rescue of cell death in *Setd2*-deficient cells by prolonged cellular aggregate formation (CAd12 days instead of CAd8, 4 days longer in retinoic acid treatment). Similar results were obtained from two independent replicates. **(B)** Shown are percentages of survival after dissociation (plated/attached) of at least two independent experiments described in (A). **(C)** CD24 and CD56-marker flow cytometry measurements in WT and *Setd2-/-* NPCs after prolonged (day 12) retinoic acid treatment. Shown are percentages of CD24-/CD56- double negative cells. **(D)** Differential expression analysis of various marker genes in WT and *Setd2-/-* NPCs by RT-qPCR (n=3). *Hprt* served as housekeeping control. Statistical significance assessed by Student's t-test ($p^{**} < 0.01$, $p^{***} < 0.0001$). **(E)** Schematic overview of the generation of an ES cell line harbouring a Tet-inducible *Neurogenin2-Neurogenin1* (i*Ngn*) transcription factor cassette in a *Setd2-/-* background. RMCE - recombination-mediated cassette exchange. **(F)** RT-qPCR analysis for *Neurog1* expression in two independent Tet-inducible i*Ngn* ES cell lines (n = 2) in comparison to the parental *Setd2-/-* only ES cell line. *Hprt* served as housekeeping control. Cells were treated with 1 µg/ml doxycycline (DOX). **(G)** Immunoblot analysis for Neurogenin 1 (NEUROG1) in two Tet-inducible *Neurogenin2-Neurogenin1* (i*Ngn*) ES cells in comparison to the parental *Setd2-/-* cell line. LAMIN B served as loading control. Cells were treated either without or continuously with 1 µg/ml doxycycline (DOX) during *in vitro* differentiation.

Supplementary Figures S9



Supplementary Figure S9 - Changes in the protein-chromatin interactome upon H3K36me3 loss.

(A) Stable isotope labelling with amino acids in cell culture (SILAC) and subsequent mass spectrometry detection of chromatin-bound proteins in WT and *Setd2-/-* mES cells. grey dots - nonspecific protein interactions, blue dots - significant only in one experiment, red dots - significant in both experiments. **(B)** Immunoblotting for Ser2-P Pol II and PCNA (loading control) of chromatin-bound fractions in wild-type and *Setd2-/-* mES cells. **(C)** Quantification of unique MS peptide counts obtained using the SETD2-SRI domain engineered chromatin reader-BASU fusion in comparison to a reader-free nuclear BASU control in WT and *Setd2-/-* ES cell backgrounds. **(D)** Venn diagram showing the overlap and difference between significantly enriched proteins from SRI-BASU WT and *Setd2-/-* mES cells. **(E)** Gene ontology (PantherGO) enrichment analysis for all differentially enriched proteins in SETD2-SRI domain engineered chromatin reader-BASU fusion in *Setd2*-deficient mES cells.

Table 1 - Differentially expressed genes in different mES cell lines from single-cell RNA-seq data of ~300 cells each (p-value < 0.01, log2FC > I0.4I).

***Setd2-/-* vs. WT**	**log2FC**	**pct.1**	**pct.2**
Grk4--chr5	-1.2	0.95	0.99
Kcnq1ot1--chr7	-1.19	1	1
Gm17821--chr12	-1.12	0.99	1
Vmn2r55--chr7	-1.04	0.97	0.98
A630089N07Rik--chr16	-1	0.99	1
Zfp71-rs1--chr13	-0.54	0.79	0.83
Erdr1--chrY	-0.46	0.96	0.99
Upp1--chr11	0.41	0.97	0.9
8430410A17Rik--chr6	0.43	0.95	0.79
Bnip3--chr7	0.51	0.88	0.75

***Eed-/-* vs. WT**	**log2FC**	**pct.1**	**pct.2**
Ifitm1--chr7	-0.94	0.86	0.98
Apoe--chr7	-0.85	0.92	0.98
Gpx4--chr10	-0.74	1	0.99
Mycn--chr12	-0.73	0.87	0.97
Mif--chr10	-0.7	1	1
Fn1--chr1	-0.69	0.96	1
Chchd10--chr10	-0.68	0.98	1
Ldhb--chr6	-0.64	0.8	0.91
Foxp1--chr6	-0.62	0.54	0.91
Gm10664--chr8	-0.59	0.89	0.97
Eed--chr7	-0.58	0.7	0.94
Usp28--chr9	-0.58	0.65	0.93
Map1b--chr13	-0.58	0.77	0.92
Trh--chr6	-0.58	0.62	0.84
Tdrd12--chr7	-0.57	0.51	0.87
Sars--chr3	-0.57	0.9	0.96
Tpi1--chr6	-0.55	0.83	0.93
Pgk1--chrX	-0.54	1	1
Gab1--chr8	-0.52	0.85	0.98
Ifitm3--chr7	-0.52	0.44	0.7
Lima1--chr15	-0.51	0.41	0.84
Ftl1--chr7	-0.5	1	1
Bnip3--chr7	-0.5	0.42	0.75
Rn45s--chr17	-0.48	1	1
Utrn--chr10	-0.47	0.62	0.93
Dhx16--chr17	-0.47	0.82	0.95
Fxyd6--chr9	-0.47	0.66	0.87
Malat1--chr19	-0.47	0.94	0.98
ERCC-00113	-0.46	0.99	1
Pdk1--chr2	-0.45	0.55	0.89
Psat1--chr19	-0.44	1	1
Pla2g1b--chr5	-0.44	0.83	0.94
Nedd4--chr9	-0.43	1	1
Fam60a--chr6	-0.43	0.97	0.99
Phc1--chr6	-0.43	0.92	0.98
Plac8--chr5	-0.43	0.65	0.81
Rbpms2--chr9	-0.42	0.95	0.99
ERCC-00096	-0.42	1	1
Fabp3--chr4	-0.42	0.6	0.78
Gpt2--chr8	-0.41	0.43	0.81
Emb--chr13	-0.41	0.89	0.97
Alg13--chrX	-0.4	0.93	0.96
Tdh--chr14	-0.4	0.98	0.99
Aldoa--chr7	-0.4	1	1
Acaa2--chr18	0.4	0.82	0.4

TKO vs. WT	**log2FC**	**pct.1**	**pct.2**
Erdr1--chrY	-0.96	0.7	0.99
Dnmt3a--chr12	-0.85	0.61	0.95
Ifitm1--chr7	-0.73	0.92	0.98
Dnmt3b--chr2	-0.7	0.54	0.9
Plac8--chr5	-0.67	0.21	0.81
Map1b--chr13	-0.63	0.65	0.92
Dnmt1--chr9	-0.62	0.51	0.95
Apoe--chr7	-0.57	0.92	0.98
Grb10--chr11	-0.56	0.45	0.82
Phlda2--chr7	-0.55	0.78	0.94
Rn45s--chr17	-0.55	1	1
Ftl1--chr7	-0.53	1	1
Tdrd12--chr7	-0.51	0.48	0.87
Gm10664--chr8	-0.5	0.85	0.97
Mir5109--chr5	-0.5	0.85	0.89
Fxyd6--chr9	-0.47	0.63	0.87
Eef2--chr10	-0.47	1	1
Malat1--chr19	-0.43	0.93	0.98
Eif2s3y--chrY	-0.42	0.05	0.49
Mif--chr10	-0.4	1	1
Aldoa--chr7	-0.4	1	1
Mirg--chr12	0.4	0.65	0.03
Ube2a--chrX	0.4	0.99	0.9
Rtp3--chr9	0.41	0.87	0.62
Marcksl1--chr4	0.41	1	1
mtNd3--chrM	0.41	1	1
Sub1--chr15	0.42	1	1
Sparc--chr11	0.43	0.99	0.96
Adprh--chr16	0.45	0.95	0.9
Jarid2--chr13	0.45	1	0.99
Dazl--chr17	0.46	0.6	0.23
Peg10--chr6	0.49	0.8	0.48
Ccrn4l--chr3	0.49	0.95	0.88
Aass--chr6	0.5	0.65	0.26
Mt1--chr8	0.56	1	0.98
Rhox5--chrX	0.57	0.81	0.43
Gabarapl2--chr8	0.59	0.97	0.94
Khdc3--chr9	0.6	0.9	0.56
Mt2--chr8	0.61	1	0.97
Rpl39l--chr16	0.64	0.95	0.71
Mest--chr6	0.65	0.97	0.85
Hspb1--chr5	0.68	1	0.98
Meg3--chr12	1.52	0.99	0.15

QKO vs. WT	**log2FC**	**pct.1**	**pct.2**
Dnmt3a--chr12	-0.85	0.63	0.95
Apoe--chr7	-0.75	0.9	0.98
Dnmt3b--chr2	-0.73	0.56	0.9
Dnmt1--chr9	-0.63	0.47	0.95
Plac8--chr5	-0.63	0.32	0.81
Gpx4--chr10	-0.57	0.99	0.99
Tpi1--chr6	-0.55	0.79	0.93
Map1b--chr13	-0.53	0.76	0.92
Fabp3--chr4	-0.53	0.48	0.78
Grb10--chr11	-0.52	0.52	0.82
Ftl1--chr7	-0.5	1	1
Ldhb--chr6	-0.49	0.81	0.91

Dcaf12l1--chrX	0.4	0.79	0.35
Ddah1--chr3	0.4	0.91	0.6
Ciapin1--chr8	0.41	0.99	0.98
Ass1--chr2	0.41	0.9	0.7
Basp1--chr15	0.41	0.96	0.91
Krt8--chr15	0.41	0.52	0.23
Tex13--chrX	0.42	0.63	0.03
Bzw1--chr1	0.43	0.99	0.97
Ckap4--chr10	0.43	0.93	0.76
Tubb3--chr8	0.44	0.94	0.87
Hmgn2--chr4	0.45	1	0.99
Peg10--chr6	0.45	0.84	0.48
Rtp3--chr9	0.45	0.82	0.62
Mylpf--chr7	0.46	0.92	0.68
Cdc5l--chr17	0.47	0.99	0.97
Khdc3--chr9	0.47	0.93	0.56
Gjb3--chr4	0.47	0.91	0.61
Pcsk6--chr7	0.48	0.92	0.45
Bmp4--chr14	0.48	0.76	0.39
Pmaip1--chr18	0.48	0.82	0.58
Jam2--chr16	0.49	0.9	0.64
Gpx1--chr9	0.49	1	0.99
Arl6ip1--chr7	0.51	1	0.97
Tex19.1--chr11	0.51	0.92	0.81
Id1--chr2	0.51	0.83	0.74
Col4a1--chr8	0.52	0.72	0.14
Llgl2--chr11	0.52	0.95	0.8
Atp5g1--chr11	0.53	1	0.99
Klf4--chr4	0.53	0.95	0.78
S100a10--chr3	0.54	0.98	0.81
Hmga2--chr10	0.55	0.87	0.48
Cycs--chr6	0.57	1	1
S100a6--chr3	0.58	0.96	0.71
Tbx3--chr5	0.59	0.88	0.28
H19--chr7	0.6	0.67	0.48
Dusp9--chrX	0.62	0.95	0.73
E130012A19Rik--chr11	0.62	0.97	0.89
Mt2--chr8	0.65	1	0.97
Slc25a4--chr8	0.66	1	0.97
Sparc--chr11	0.68	0.99	0.96
Lefty2--chr1	0.7	0.77	0.6
Hspb1--chr5	0.73	1	0.98
Calcoco2--chr11	0.83	0.92	0.5
Gabarapl2--chr8	1.02	1	0.94

Chchd10--chr10	-0.48	0.98	1
Phlda2--chr7	-0.46	0.81	0.94
Ifitm1--chr7	-0.45	0.96	0.98
Fxyd6--chr9	-0.43	0.61	0.87
Eif2s3y--chrY	-0.42	0.06	0.49
Sars--chr3	-0.42	0.89	0.96
Pgk1--chrX	-0.4	1	1
Gm10664--chr8	-0.4	0.9	0.97
S100a10--chr3	0.4	0.95	0.81
Rhox13--chrX	0.41	0.72	0.32
Efhc2--chrX	0.42	0.72	0.07
Gm13247--chr4	0.42	0.87	0.52
Tubb3--chr8	0.42	0.95	0.87
mtNd3--chrM	0.42	1	1
Syce1--chr7	0.43	0.87	0.48
Dusp9--chrX	0.44	0.92	0.73
Cenpm--chr15	0.44	0.95	0.9
Mirg--chr12	0.46	0.74	0.03
Vim--chr2	0.47	0.88	0.7
Ybey--chr10	0.48	0.9	0.69
Gabarapl2--chr8	0.5	0.99	0.94
Stmn2--chr3	0.51	1	0.98
Hspb1--chr5	0.51	1	0.98
1700013H16Rik--chrX	0.54	0.54	0.04
Rtp3--chr9	0.56	0.85	0.62
Mt1--chr8	0.56	0.99	0.98
Gpx1--chr9	0.57	1	0.99
Sparc--chr11	0.61	1	0.96
Mest--chr6	0.62	0.96	0.85
Rpl39l--chr16	0.74	0.96	0.71
Mt2--chr8	0.75	1	0.97
Dazl--chr17	0.81	0.8	0.23
Calcoco2--chr11	0.82	0.83	0.5
Meg3--chr12	1.69	0.98	0.15

Table 2 - Cell Lines.

Cell Line	Source
RMCE-competent, BirA+ mES cells	Baubec, et al., 2013
Dnmt1, *Dnmt3a*, *Dnmt3b-/-* (TKO) mES cells	Domcke, et al. 2015
Eed-/- mES cells	Villasenor, et al., 2020
Setd2-/- mES cells	Baubec, et al., 2015
Setd2-/- in TKO background mES cells	generated
constitutive sh*Setd2* mES cells	generated
Tet-inducible sh*Setd2* mES cells	generated
Tet-inducible *Ngn2-Ngn1* mES cells	generated
Tet-inducible sh*Exosc3* mES cells	generated
Mettl14-/- mES cells	generated
BASU-SRI WT mES cells	generated
BASU-SRImut *Setd2-/-* mES cells	generated
Setd2 endogenously tagged C-terminal FLAG-bio mES cells	generated
Setd2-/- SETmut addback mES cells	generated

Table 3 - Oligo Sequences.

Target	Sequence 5'-3'
Setd2 -/- exon 3 sgRNA 5'	GTCGGTCCGAAAGAGATCGA
Setd2 -/- exon 3 sgRNA 3'	AGCGTGTCCTCTCACGATAA
Setd2 -/- genotyping	GGGGAAATCATCAAGATCGAA
Setd2 -/- genotyping	TTAGGTCTCTGTAGGAATGGG
Setd2 shRNA 1	ATAGTGTGACCTCGCCTTATT
Setd2 shRNA 2	ACTTTGTGAGGATAGTATAAA
RT-qPCR primer *Setd2* forward	TGGGGCCTTCGTGTGCTATG
RT-qPCR primer *Setd2* forward	GCAATCTTCTCCACATGCTACTT
RT-qPCR primer *Hprt* forward	GGACCTCTCGAAGTGTTGGA
RT-qPCR primer *Hprt* reverse	CAACTTGCGCTCATCTTAGGC
RT-qPCR primer *Oct4* forward	AAGCGAACTAGCATTGAGAACC
RT-qPCR primer *Oct4* reverse	CATACTCGAACCACATCCTTCTC
RT-qPCR primer *Nanog* forward	GAACTCTCCTCCATTCTGAACCT
RT-qPCR primer *Nanog* reverse	GACCATTGCTAGTCTTCAACCAC
RT-qPCR primer *Sox2* forward	ACAGCATGTCCTACTCGCAG
RT-qPCR primer *Sox2* reverse	ATGCTGATCATGTCCCGGAG
RT-qPCR primer *Klf4* forward	GCAGTCACAAGTCCCCTCTC
RT-qPCR primer *Klf4* reverse	TTTGCCACAGCCTGCATAGT
RT-qPCR primer *Gata6* forward	ACAGTCCCCGTTCTTTTACTGAG
RT-qPCR primer *Gata6* reverse	GGTACAGGCGTCAAGAGTGTTAC
RT-qPCR primer *Fgf4* forward	CTACCGGATAGGAGACCCTTAGA
RT-qPCR primer *Fgf4* reverse	CCTCTTGTCCATTCCTAGTTCCT
RT-qPCR primer *Gata4* forward	TGGGGAGATTAGGTGAGGGG
RT-qPCR primer *Gata4* reverse	ATTAGCTGCACAACTGGGCT
RT-qPCR primer *Cdx2* forward	CTTTGTCAGTCCTCCGCAGT
RT-qPCR primer *Cdx2* reverse	CATTGAGACCGTGGGCTACC
RT-qPCR primer *Sox17* forward	ATGCATTCTGGACCCGCTAC
RT-qPCR primer *Sox17* reverse	AGCTCTCCTGCCTCTCAGAA

Target	Sequence 5'-3'
RT-qPCR primer *Nestin* forward	CCTTGCCTAATACCCTTGAGACT
RT-qPCR primer *Nestin* reverse	TTTTAGGATAGGGAGCCTCAGAC
RT-qPCR primer *Olig2* forward	TGACTCCCTGTCTGGGCTTA
RT-qPCR primer *Olig2* reverse	AGTGCTTCTGATACCCACGC
RT-qPCR primer *Neurod1* forward	CCCTACTCCTACCAGTCCCC
RT-qPCR primer *Neurod1* reverse	GAGGGGTCCGTCAAAGGAAG
RT-qPCR primer *Pax3* forward	CATGCCCGGGTTCTCTCTTT
RT-qPCR primer *Pax3* reverse	GTCCCATGGTTGCGTCTCTA
RT-qPCR primer *Pax6* forward	GCCACCAGACTCACCTGACACC
RT-qPCR primer *Pax6* reverse	CTCACCGCCCTTGGTTAAAGTC
RT-qPCR primer *Crabp1* forward	AACTTCAAGGTCGGAGAGGG
RT-qPCR primer *Crabp1* reverse	GCTCTCGGGTCCAGTAAGTT
RT-qPCR primer *Onecut2* forward	CGGGCCATGACAAAATGCTC
RT-qPCR primer *Onecut2* reverse	CCAGGGTGGTGTAAGCCATT
RT-qPCR primer *Foxd3* forward	GCAACTACTGGACCCTGGAC
RT-qPCR primer *Foxd3* reverse	GCTCCGAAGCTCTGCATCAT
RT-qPCR primer *Neurog1* forward	ATTACATCTGGGCGCTCACC
RT-qPCR primer *Neurog1* reverse	GAAGGTGGAGAAGGGCTGTC
RT-qPCR primer *Sox4* forward	CACAACGCCGAGATCTCCAA
RT-qPCR primer *Sox4* reverse	CCGACTTCACCTTCTTTCGC
RT-qPCR primer *Msx3* forward	GAGTGCGCGACTGGAGG
RT-qPCR primer *Msx3* reverse	CACAGAGCACGGACCACTC
RT-qPCR primer *Epha3* forward	TTGGGGACTTGCAAGGAGAC
RT-qPCR primer *Epha3* reverse	TTCAGAATGCGATCCCCGAG
RT-qPCR primer *Zic1* forward	TTGAAAGCAGCGCTGGAGTA
RT-qPCR primer *Zic1* reverse	CCTTCAAGCTCAACCCCAGT
Setd2-SRI domain cDNA forward	ACAGCAGAAGCAGACACCTC
Setd2-SRI domain cDNA reverse	CTCTAGTTCAGTGTCCTCTT

Table 4 - Omics data sets.

ChIP-seq
ES WT H3K27me3 (2 replicates)
ES WT H3K4me3 (2 replicates)
ES WT H3K9me3
ES WT H3K36me3 (2 replicates)
ES WT total Pol II (2 replicates)
ES WT Ser2-P Pol II (2 replicates)
ES WT H4 panacetyl (2 replicates)
ES *Setd2-/-* H3K27me3 (2 replicates)
ES *Setd2-/-* H3K4me3 (2 replicates)
ES *Setd2-/-* H3K9me3
ES *Setd2-/-* H3K36me3 (2 replicates)
ES *Setd2-/-* total Pol II (2 replicates)
ES *Setd2-/-* Ser2-P Pol II (2 replicates)
ES *Setd2-/-* H4 panacetyl (2 replicates)
NP WT H3K27me3 (2 replicates)
NP WT H3K4me3 (2 replicates)
NP WT H3K36me3 (2 replicates)
NP WT total Pol II (2 replicates)
NP WT Ser2-P Pol II (2 replicates)
NP WT H4 panacetyl (2 replicates)
NP *Setd2-/-* H3K27me3 (2 replicates)
NP *Setd2-/-* H3K4me3 (2 replicates)
NP *Setd2-/-* H3K36me3 (2 replicates)
NP *Setd2-/-* total Pol II (2 replicates)
NP *Setd2-/-* Ser2-P Pol II (2 replicates)
NP *Setd2-/-* H4 panacetyl (2 replicates)
TNd6 H3K36me3 + Active Motif Spike-In WT (2 replicates)
TNd6 H3K36me3 + Active Motif Spike-In *Setd2-/-*
TNd6 H3K36me3 + Active Motif Spike-In shSetd2 -DOX
TNd6 H3K36me3 + Active Motif Spike-In shSetd2 +DOX

ATAC-seq
ES WT
ES *Setd2-/-*
NPC WT
NPC *Setd2-/-*
RNA-seq
ES scRNA WT
ES scRNA *Setd2-/-*
ES scRNA *Eed-/-*
ES scRNA TKO
ES scRNA *Setd2-/-* in QKO
NP scRNA WT
NP scRNA *Setd2-/-*
ES PolyA-RNA WT
ES PolyA-RNA *Setd2-/-*
NP PolyA-RNA WT
NP PolyA-RNA *Setd2-/-*
TN PolyA-RNA WT
Mass Spectrometry
ES SILAC WT vs *Setd2-/-*
ES Ser2-P Pol II ChIP WT
ES Ser2-P Pol II ChIP *Setd2-/-*
NP Ser2-P Pol II ChIP WT
NP Ser2-P Pol II ChIP *Setd2-/-*
ES SRI-BASU WT
ES SRI-BASU *Setd2-/-*
ES empty-BASU WT
ES empty-BASU *Setd2-/-*
Nascent RNA-seq
ES PRO-seq WT
ES PRO-seq *Setd2-/-*

5. Discussion

Here, we addressed the contribution of chromatin modifications to gene expression noise and cellular identity. We further validated the essentiality of these different chromatin modifications to the establishment of new gene expression programs assayed by the differentiation of respective knock-out embryonic stem cell lines to neuroectoderm lineages using a well-established neurodifferentiation protocol (Bibel et al., 2007). We were able to identify a novel role of SETD2 in the establishment of correct transcriptional profiles upon terminal differentiation. Additional experiments indicate that this could be potentially influenced by altered interactions with the co-transcriptional splicing machinery.

We first confirmed that H3K27me3, H3K36me3 or DNA methylation are dispensable for the survival and maintenance of pluripotency of ES cells (Lei et al., 1996; OGEE database, Chamberlain et al., 2008; Zhang et al., 2014). We could further validate that chromatin modifications like H3K27me3 and H3K36me3 are predictive of gene expression noise (Faure et al., 2017) in ES cells (Figure 1). However, they do not contribute to a cell-to-cell variation in ES cells, since their targeted depletion did not show additional differences. It is already well established that ES cells are highly amendable to changes. Thus, adaptation to the loss of chromatin modifications could be a result of their plasticity in the pluripotent state.

Another reason for the lack of transcriptional changes in ES cells, could be the redundancy between different chromatin marks. For instance, DNA methylation loss is important to silence repetitive elements, but is dispensable in ES cells due to the additional presence of H3K9me3 at these sites (Rowe et al., 2010; Karimi et al., 2011). Moreover, H3K36me3 has been recently shown to maintain levels of Disruptor of telomeric silencing 1-like (DOT1L)-mediated H3K79me2 on target genes critical for cancer cell survival (Skucha et al., 2018). Thus, it will be necessary to elucidate such a combinatorial effect or interplay between chromatin factors and marks on cell survival in the pluripotent state or during differentiation in more detail. For this, CRISPR-Cas9 screens could be utilised to induce synthetic lethality in mES cells lacking different chromatin marks or to identify factors which promote cell survival in these compromised cells during differentiation. Positive hits could then be characterised in regards to their mechanism of action which might give insights into potential therapeutical advances, too.

In contrast to the lack of changes in the pluripotent state, differentiation experiments to neuroectoderm showed that loss of these modifications does influence development in a critical manner (Figure 2). Consequences of H3K27me3 and DNA methylation loss established early after exit from pluripotency, when silencing of developmental genes becomes crucial for the establishment of correct gene expression programs (Bracken et al., 2006; Ambrosi et al., 2017). TKO/QKO cells might be further compromised due to the upregulation of repetitive elements upon exit of pluripotency which can potentially harm genome stability (Walsh et al., 1998). In contrast, loss of the H3K36me3 mark was dispensable for early neuronal development stages. Only upon terminal differentiation, *Setd2-/-* cell survival was dramatically compromised in agreement with the *Setd2-/- in vivo* phenotype, showing lethality and neuronal tube closure defects at E10.5 (Hu et al., 2010).

With the help of inducible knockdown experiments, we could further elucidate that SETD2/ H3K36me3 is necessary for lineage commitment of neuronal precursors to post-mitotic neurons, but its loss does not impair lineage maintenance. This goes in line with the differential expression of *Setd2* during mouse development and in adult tissues, as its highest expression is present during embryonic development in comparison to adult tissues as seen from mouse ENCODE transcriptome data (Yue et al., 2014). Furthermore, other studies have shown that loss of SETD2 decreases the differentiation and terminal commitment potential to lineages of other germ-layers such as the primitive endoderm (Zhang et al., 2014), implying that the loss of H3K36me3 generally compromises terminal differentiation in a non-lineage specific fashion.

We were able to identify a change in transcriptional rewiring in *Setd2-/-* neuronal progenitors as the cause for the observed differentiation failure (Figure 3). Changes in transcription of *Setd2-/-* neuronal progenitors were minor as seen in other cellular models (Ho et al., 2016; Park et al., 2016a; Tiedemann et al., 2016), but most differentially expressed genes indicated an impaired neuronal gene expression network. *Setd2-/-* NPCs also failed to control expression of other lineage marker or to completely downregulate pluripotency markers. However, this change in bulk gene expression could not be explained by increased transcriptional noise in single *Setd2-/-* NPCs, as it has been recently suggested based on the association of H3K36me3 with genes of low cell-to-cell variation in mES cells (Faure et al., 2017).

Instead, we suggest that the reduced differentiation into final lineages could partially caused by failure to establish accurate gene expression patterns in absence of SETD2. We were further able to partially bypass this inaccuracy by overexpression of a major transcriptional driver for neuronal

lineage determination (Figure 4). Its induced expression moderately reverted incorrect gene expression in the H3K36me3-deficient neuroprogenitors and let to an increased in survival. Nonetheless, we only investigated a subset of genes changed in the rescued cells which might not be representative for all transcriptome changes. Thus, additional bulk RNA-seq experiments of mutant cells expressing an induced neuronal transcription factor will shed light on an overall rescue of the SETD2/H3K36me3 deficiency.

Since the correction of gene expression did not lead to a full reversion of the differentiation phenotype, we need to consider that there are other mechanisms in place which fail to operate correctly in the absence of SETD2. For instance, m6A on longer exons or towards the 3' end of the gene correlates with the accumulation of H3K36me3 on actively transcribed genes in mES cells (Huang et al., 2019). Loss of H3K36me3 potentially leads to an impaired recruitment of the m6A machinery to the nascent mRNA. Reduced m6A levels can have various effects on mRNA processing such as a stabilisation of target transcripts or a failure in mRNA splicing and export, potentially influencing protein abundance (see chapter 1.1.4.3.). These effects could provide explanations for the differential expression in *Setd2-/-* NPCs, such as the insufficient downregulation of pluripotency marker genes. Thus, measuring global and genome-wide m6A level changes in the knock-out cell line by mass spectrometry will give insights into a potential relationship between mRNA methylation and the observed cell death. However, the direct relationship between METTL enzymes and H3K36me3 remains to be elucidated. Additionally, we did not observe differential transcript abundance of pluripotency genes in our analyses of knock-out cells as suggested recently in mES cells (Huang et al., 2019).

Another alternative explanation might be the non-histone directed methyltransferase activity of SETD2. Neuronal progenitors have to undergo drastic morphology changes when committing to become post-mitotic neurons which is highly dependent on the precise arrangement of the cytoskeleton. SETD2 has been recently shown to methylate not only H3K36me3, but also the lysine 40 of α-tubulin (TubK40me3) (Park et al., 2016b), which is necessary for genomic stability. Thus, *Setd2*-deficient cells might not be able to adapt to the new structural requirements and induce cell death. Therefore, next steps will need to include TubK40me3 level analysis by immunoblot experiments and microtubule immunoflouresence staining in mutant neuronal progenitors and surviving post-mitotic neurons. However, since a defect in the cytoskeleton would be expected to impair all the cells to the same extent, α-tubulin methylation may not be sufficient to fully explain the remaining ~20 % surviving cells after terminal differentiation or the considerable rescue observed upon *Neurogenin 1/2* expression.

Differential gene expression can originate from many different variabilities in pathways, cues and machineries involved in the process of transcription and co-transcriptional processes. We explored several different possibilities to assess the gene expression changes upon SETD2/H3K36me3 loss in neuronal progenitor cells. First, we were not able to detect major changes in the chromatin landscape through ChIP-seq experiments of Pol II and other histone modifications, except for a known increase in acetylation upon H3K36me3 loss (Lee et al., 2013). Impaired compaction of chromatin after Pol II passage during transcription due to an hyperacetylated state can lead to an increase in spurious transcription from sites. Nonetheless, we did not observe any increase in spurious initiation, antisense transcription in absence of SETD2. Furthermore, ATAC-seq experiments did not suggest increased accessibility at transcribed gene bodies in absence of H3K36me3. However, the presence of spurious transcripts and initiation remains to be analysed using more sensitive assays such as PRO-seq or CAGE-seq (Shiraki et al., 2003).

Chromatin modifications stand in crosstalk to each other, are mutually exclusive and sometimes offer redundancy (Zhang et al., 2015). This could further explain the lack of changes in the H3K36me3-depleted chromatin environment in NPCs. Investigation of other chromatin marks need to be assessed to rule out their role in altering the transcriptional output. For instance, H3K27me1 coincides with H3K36me3 to promote active transcription (Ferrari et al., 2014). In the absence of H3K36me3, H3K27me2 accumulates instead and represses gene transcription, potentially influencing the whole transcriptome. Additionally, technical draw-backs of ChIP sequencing need to taken into consideration. For instance, Pol II ChIP-seq only captures a static view at a point in time, which might not recapitulate dynamic changes in Pol II elongation behaviour that has been shown to be influenced by H3K36me3 (Jonkers et al., 2014; Wen et al., 2014). Dynamic methods that allow for this are nuclear run-on assays such as PRO-seq that map the genome-wide distribution of transcriptionally-engaged Pol II at base-pair resolution (Jonkers and Lis, 2015). Applying this to our cell system could help to elucidate transcriptional profiles in presence and absence of H3K36me3 at high resolution by globally mapping strand-specific Pol II density.

The downstream effects of histone K36 methylation are facilitated by PWWP-domain reader proteins (Hyun et al., 2017). Thus, we decided to analyse the chromatin-protein interactome in presence or absence of SETD2/H3K36me3 using mass spectrometry. First analyses of chromatin-bound fractions in WT and *Setd2-/-* mES cells using SILAC did not offer enough resolution to identify differential protein interactions at actively transcribed gene bodies. Therefore, we performed Ser2-P Pol II ChIP-MS in mES and neuronal progenitor cells to enrich for active chromatin sites. We could indeed identify proteins known to be involved in elongation such as Pol II subunits or elongation factors as well as proteins involved co-transcriptional processes, but there

were no significant differences between wild-type and *Setd2-/-* cells. Since ChIP involves cross-linking and antibody-based enrichment, it might again not be the most optimal approach. Cross-linking of DNA and protein by formaldehyde only covers proteins in a physical range of ~2 angstrom (Hoffman et al., 2015). Thus, a pulldown for the elongating form of the multi-subunit complex Pol II might not be able to reveal changes caused by the lack histone modifications. This initial MS run will need repetition and optimisation, e.g. by increasing the range of interactions using the disuccinimidyl glutarate (DSG) cross-linker in addition to formaldehyde (Tian et al., 2012). However, passaging of Pol II might generally disrupt the chromatin environment which is usually present at H3K36me3-rich sites, rendering this method less suitable for our purpose.

In order to avoid the disadvantages mentioned above, we utilised our recently described tool ChromID (Villasenor et al., 2020). The non-redundant SETD2 binds with its SRI-domain to the Ser2-P CTD of Pol II (Kizer et al., 2005). Fusing this domain to an active BirA ligase (BASU) and analysing the biotinylated, nuclear proteins helped to understand the chromatin interactome at H3K36me3 gene bodies. We were able to identify a small number of interactions in mES cells which were significantly lost in the *Setd2-/-* background. These proteins were amongst others enriched for the process of co-transcriptional splicing, indicating that lack of H3K36me3 indeed could influence the recruitment of splicing factors to chromatin (Luco et al., 2010; Zhu et al., 2017). In order to assess to what extent these factors play a role for the impaired progression to the terminal neuronal lineage, remains to be elucidated by repetition of the SRI-BASU experiments in neuronal progenitors. If the depletion of splicing factors holds true in NPCs, these protein hits need to be further characterised in order to find the link between observed gene expression changes and depletion of factors involved in splicing.

Analysis centered on the differentially expressed genes in bulk RNA-seq data sets will also give further insights on how splicing or spurious transcription in *Setd2-/-* cells is altered. This can be extended by enriching for spurious transcripts in the cell through interference with the exosome complex, by utilising established inducible sh*Exosc3* cell lines in wild-type and *Setd2-/-* backgrounds. Splicing defects can also generally alter the mRNA abundance which can highly vary between different cells in an isogenic population (La Manno et al., 2018). Thus, to quantitatively assess an effect of H3K36me3 loss on mRNA abundance and its time-resolved changes during lineage progression, referred to as RNA velocity, the single-cell RNA-seq data sets in hand need further exploration to examine differential kinetics of transcription in *Setd2-/-* cells.

It was recently shown that H3.3K36M mutants do not show an impaired differentiation potential towards neuroectoderm (Gehre et al., 2020). Even though the H3.3 incorporation at this stage is

not complete and canonical H3 histone is still present, the successful progression in lineage commitment hints to an individual role of SETD2 independent of its histone mark. We have recently generated *Setd2-/-* ES cells expressing a catalytic dead (SET domain) mutant in order to distinguish individual roles between the methylation mark and the enzyme. However, neurodifferentiation experiments as well as transcriptome and proteome analysis of these cells are still pending.

In summary, we showed here that there is a connection between SETD2-mediated H3K36me3 and rewiring of gene expression circuits during terminal differentiation. However, it remains to be investigated how the establishment of transcriptional programs is specified by individual and combined roles of the enzyme, its mark and downstream signalling pathways. Dissecting their individual and overlapping contribution to transcription and RNA maturation could provide significant insights into gene regulation, and will help to understand processes which lead to altered gene expression in disease. Given the presence of the H3K36me3 mark on all transcribed genes of a cell, the concerted influence could potentially affect the entire transcriptional output and subsequently the whole identity of a cell. Accordingly, it is easily conceivable that analysing such factors and chromatin marks involved in cancer development with respect to their function in transcription can lead to the identification of new therapeutic targets and improved treatment.

6. Material and Methods

Culturing and differentiation of ES cells to terminal neurons

Mouse embryonic stem cells (159-2) were cultured on 0.2 % gelatine-coated dishes in DMEM (Invitrogen) supplemented with 15 % fetal calf serum (Invitrogen), 1× non-essential amino acids (Invitrogen), 1x Glutamax (Invitrogen), homemade leukemia inhibitory factor (LIF), and 0.001 % b-mercaptoethanol (Invitrogen) at 37°C in 7 % CO_2. For -LIF experiments ES cells were cultured without LIF on gelatine-coated plates. Inducible knockdown as well as *iNgn* cell lines were treated with 1 µg/ml doxycycline (Sigma Aldrich). ES cell lines were differentiated as previously described (Bibel et al), except that no feeder cells were used. Microscopy images were taken with 100 x magnification. Cell count assays were performed using live/dead stain (BioRad) and TC20™ Automated Cell Counter (BioRad). For DNA damage analysis, cells were treated for 1 h with 1µM Camptothecin (Sigma Aldrich) and processed for QIBC as described previously (Teloni et al., 2019).

Cell line generation

All generated cell lines are found in Table 2. *Setd2-/-* in *Dnmts* TKO mouse ES cell background (courtesy of Dirk Schübeler, FMI, Switzerland) was generated using CRISPR/Cas9. The sgRNAs target two parts the third exon, resulting in a deletion of approximately 200 bp. The Cas9 sgRNA sequences were cloned into the pX330-U6-Chimeric_BB-CBh-hSpCas9 (addgene 42230). Transfections together with pRR-Puro recombination reporter (Flemr and Buhler, 2015) (Addgene 65853) were conducted using Lipofectamine 3000 reagent (L3000015, Thermo Fisher Scientific) at a 2:1 Lipofectamine/DNA ratio in OptiMEM (31985070, Thermo Fisher Scientific). 36 h later cells were treated with 2 µg/ml puromycin for another 36 h. *Setd2-/-* ES cells (Baubec et al., 2015) with an inducible neurotranscription factor (i*Ngn*) were generated as described previously with adaptations (Busskamp et al., 2014). Doxycline-inducible rtTA3 system (addgene 61472) was randomly integrated into *Setd2-/-* ES cells (Baubec et al., 2015) using 20 µg of linearised plasmid with bleomycin resistance. Cells were treated with 200 µg/ml Zeocin™ (InvivoGen). The TetON-inducible *Ngn2*-2A-*Ngn1 Setd2-/-* ES cells were then obtained by recombinase-mediated cassette exchange (RMCE) (Baubec et al., 2013), based on transfection of RMCE constructs (addgene 61471 into pCAGbio_V6) with an expression plasmid for CRE recombinase in a 1:0.6 DNA ratio. Similarly, shRNA-inducible ES cells were generated using a TetON-inducible system for a shRNA,

expressed from the RMCE site, against *Setd2*. For ChromID cell line generation, BASU or TurboID-linked SRI domain was cloned into RMCE targeting vector parbit-v9 and SRI-ChromID ES cells in a wild-type as well as *Setd2-/-* background were generated as described previously (Villasenor et al., 2020). Positive clones were confirmed by Sanger sequencing, RT-qPCR and immunoblotting. Cloning strategies are available on request. Single guide RNA, shRNA and oligo sequences are collected in Table 3.

Immunoblotting

Crude nuclear extracts cells were obtained as described in (Manzo et al., 2018), histones were extracted acid extracted according to (Villasenor et al., 2020) and chromatin-bound fractions (chromatin enrichment for proteomics, ChEP) were obtained according to (Kustatscher et al., 2014). Membranes were blocked with 5 % (wt/vol) milk in TBS-0.1 % Tween20 and incubated with primary antibodies against SETD2 (1:1000, A11271, ABclonal), Pol II (MABI0601, MBL), PCNA (sc-56, Santa Cruz Biotechnology), Phospho-H2A.X (Ser-139, 1:8000, #9718, Cell Signaling), H3K36me3 (1:5000, ab9050, abcam), anti-histone H1 (1:5000, 05-457, Millipore), anti-NEUROG1 (1:1000, sc-100332, Santa Cruz Biotechnology), or anti-LAMIN B1 (1:1000, sc-374015, Santa Cruz Biotechnology) overnight at 4°C. Protein detection was facilitated using species-specific antibodies conjugated to horseradish peroxidase and Pierce® Peroxidase IHC Detection Kit (Thermo Scientific).

RNA isolation, cDNA synthesis and RT-qPCR

RNA was isolated from mouse ES cells, cellular aggregates at day 4 and NPCs using the RNeasy Plus mini kit (Qiagen). Coding DNA was synthesised from 2 µg isolated RNA with SuperScript III First-Strand Synthesis (Invitrogen) for 60 min at 50°C using Oligo$(dT)_{18}$ (Thermo Fisher), followed by heat-inactivation for 10 min at 85°C. Residual RNA was digested with 2 units RNaseH (NEB) for 20 min at 37°C. Target sequences were quantified by real-time qPCR analysis using a KAPA SYBR FAST qPCR Kit (Sigma Aldrich) on a QuantStudio 5 Real-Time PCR System (Applied Biosystems). Comparative quantification (ddCt) was used to determine transcript levels relative to the housekeeping gene *Hprt*. Each sample was at least run in technical triplicates. Oligo sequences available in Table 3.

Surface marker and cell cycle analysis using flow cytometry

For CD24 and CD56 measurements in mES cells and neuronal progenitors, single-cell suspensions were obtained through trypsinization and filtered through 40-um cell strainers (BD Biosciences). For cell surface marker analysis, cells were stained for 30 min at 4° C with saturating concentration of anti-CD24a and anti-CD56 monoclonal antibodies in the presence of anti-CD16/

CD32 (eBioscience). For live dead cell exclusion, LIVE/DEAD Fixable Near-IR Dead Cell Stain (Invitrogen) was used. For cell cycle and Ki-67 measurements, single-cell suspensions were fixed and permeabilised for 30 min at 4°C with Foxp3/Transcription Factor Staining Buffer set (eBioscience). Anti-Ki-67 or isotype control (eBioscience) was added and incubated for 45 min at RT in permeabilisation buffer. DAPI (5 µg/ml, Sigma) was added as a fluorescent DNA stain 5 min prior to FACS measurements and incubated at RT in the dark. Samples were acquired and data was analysed as in (Villasenor et al., 2020).

Chromatin immunoprecipitation (ChIP), -sequencing and read processing

Histone ChIP experiments and sequencing were performed as previously described (Villasenor et al., 2020). Here, 100 µg chromatin were incubated with 5 µg of either total Pol II (MABI0601, MBL), Ser2-P RNA Pol II (ab5095, abcam), H3K36me3 (ab9050, abcam), H3K4me3 (ab8580, abcam), H4-panacetyl (B_2687872, ActiveMotif), H3K27me3 (C15410195, diagenode) antibody. Wiggle tracks were obtained with QuasR and visualised with the UCSC genome browser (https://genome.ucsc.edu). For quantitative ChIP-Rx sequencing experiments, the Spike-in antibody (61686) and *Drosophila melanogaster* chromatin (53083) from Active Motif were used according to manufacturer's instructions.

Genomic coordinates and analysis at gene bodies

Analysis were performed as described previously (Teloni et al., 2019), using the *Mus musculus* assembly version NCBI Build 37 mm9 from July 2007 and RefSeq gene predictions at NCBI from the annotation release: GCF_000001635.18. For genomic analyses, read counts overlapping with gene bodies (+2 kb from TSS) were used.

Assay for Transposase-Accessible Chromatin (ATAC)-sequencing

ATAC-seq experiments were performed in biological duplicates using wild-type and *Setd2-/-*mouse embryonic stem cells and neuronal progenitors. The tagmentation reaction was performed with 50,000 cells as previously described (Buenrostro et al., 2013) with minor adjustments of the protocol, using the Nextera DNA Library Prep Kit (Illumina) together with the barcoded primers from the Nextera Index Kit (Illumina). In brief, an additional size selection step was performed after the first 5 cycles of library amplification. For this, the PCR reaction was incubated with 0.6 x volume of Ampure XP beads (Beckman Coulter) for 5 min to allow binding of high molecular weight fragments. Beads containing long DNA fragments were separated on a magnet and the supernatant containing only small DNA fragments below roughly 800 bp were cleaned up using MinElute PCR purification columns (Qiagen). All libraries were amplified for 12 cycles in total, visualised and quantified with a TapeStation2200 (Agilent), and sequenced as mentioned above.

ATAC-seq data processing and differential analysis

Raw sequencing reads were processed as discussed in (Villasenor et al., 2020) (TrimGalore, QuasR, Bowtie1, mm9, PCR duplicate removal). To call differentially open chromatin sites between the various conditions, the mapped reads of all conditions were merged into one alignment file. Peaks were called on the merged sample using MACS2 with --nomodel --shift -100 --extsize 200 --keep-dup all, to obtain a consensus peak set with equal contribution of each sample independent of the respective biological treatment (Lun and Smyth, 2014). The combined peak set was used to build a count matrix using the aligned reads of the individual samples. Differential peaks between sample groups were identified using edgeR (Robinson et al., 2010). For this, the count matrix was filtered for peaks with low coverage across all samples based on a average CPM value < -1, resulting in a final set of 131271 peaks with p-values below 0.001266164. The remaining counts were normalised using the total library size as well as edgeR's TMM derived normalisation factors. Between two sample groups, only regions with a log fold change above > 0.5 and a FDR value < 0.05 were considered as differentially accessible for further analysis.

PolyA RNA-sequencing and differential gene expression analysis

Total RNA was isolated from NPCs using the RNeasy Plus mini kit (Qiagen). RNA integrity was measured using a model 2100 Bioanalyzer (Agilent). PolyA-tailed mRNAs were isolated and enriched using NEB Next Poly(A) mRNA Magnetic Isolation Module according to manufacturer's instructions. Libraries for 1 µg mRNA were prepared using NEB Next UltraTM II Directional RNA Library Prep Kit for Illumina. Sequencing of library pools and read processing were performed as described above. Read counts per gene were generated using STAR (Dobin et al., 2013) based on the gene transcript annotation gencode.mouse.v1.annotation.gtf (NCBIM37, mm9). Differential gene expression was performed using the DESeq2 package with significance set to p-value < 0.05 and log fold change > I1I (Love et al., 2014). Gene ontology enrichment analysis on differentially expressed was performed using Protein Analysis Through Evolutionary Relationships (PANTHER) classifications (Thomas et al., 2003).

Single-cell RNA-sequencing and processing

Dissociated ES cells and NPCs were sorted using a FACS Aria III cell sorter (BD Bioscience) with gates set based on size and forward scattering to exclude debris and doublets. The SORT-seq single-cell RNA-sequencing protocol was carried out as described previously (Hashimshony et al., 2016; Muraro et al., 2016). Single cells were sorted in four 384-well plates (Biorad) provided by *Single Cell* Discoveries, containing 5 µl of CEL-Seq2 primer solution in mineral oil, further containing 24 bp polyT stretch, a 4 bp random molecular barcode (UMI), a cell-specific barcode,

the 5'Illumina TruSeq small RNA kit adaptor and a T7 promoter. After sorting, the plates were immediately placed on dry-ice and stored at -80°C until sent for sequencing. SORT-seq data from Single Cell Discoveries were delivered as tables with estimated original transcript counts by a Poisson correction (Grun et al., 2014). Cells were quality checked using scater v1.12.2 (McCarthy et al., 2017). Cells fulfilling the following attributes were subjected to downstream analysis: deviating less than three NMADS from the total reads, total features, and mitochondrial reads distributions; and on top of that, manually gated to have less than 15 % mitochondrial reads and more than 2000 features. Cell cycle phase was assigned using scran version 1.10.1 `cyclone` function with default parameters (Scialdone et al., 2015). Seurat v3.0.2 was used to normalise read counts, regress out the library size, the mitochondrial reads proportion per cell and cell cycle phase, find variable features with the `vst` method, integrate multiple plates as in, dimensionality reduce the data (PCA, t-SNE, UMAP), cluster cells and test for differentially expressed genes (Stuart et al., 2019).

Variability analysis

Variability analysis was carried out using the PAGODA framework implemented within the `scde` package v1.99.1 (Fan et al., 2016). Briefly, we computed cell-specific error models and estimated the residual gene expression variance while attributing an adjusted (normalised) variability value to each gene within each genotype. This adjusted variability leverages the technical and intrinsic biological components of the variation, providing a score that is comparable across genes and independent from the gene expression level. Gene expression variability analysis integration with chromatin modifications data was carried out using normalised ChIP-seq data (Table 4) summarised by gene for signals at the TSS (+/- 2 kb from the transcriptional start site) and in gene bodies (+2 kb from transcriptional start site). ChIP-seq signals were quantilised into deciles to plot trends of normalised gene expression variability. Data analysis was carried out in R v3.6.1. Source code for analysis is available upon request.

SRI-ChromID and ChIP-MS sample preparation

ChIP-MS samples were obtained by chromatin immunoprecipitation as described above. 150 µg chromatin from wild-type and Setd2 -/- ES cells and NPCs and 7.5 µg Ser2-P Pol II antibody (abcam) were used for the each IP. Instead of de-crosslinking, protein were eluted by boiling Protein-A beads in Laemmli buffer for 20 min. Proteins were run on a NuPAGE-Novex Bis-Tris 4-12 % (Invitrogen), stained with InstantBlue™ Ultrafast Protein Stain (Sigma Aldrich) and each sample lane was cut into six equally-sized fractions excluding the antibody bands. Peptides were extracted according to (Rosenfeld et al., 1992). ChromID samples were prepared as previously described (Villasenor et al., 2020). In brief, nuclear extraction of SRI-ChromID mES cells of WT and *Setd2-/-*

background was performed after biotin incubation for 12h (BASU) and 1h (TurboID), followed by affinity purification through streptavidin beads, high-stringency washes and on-bead digestion.

Label-free MS data acquisition and analysis

MS data acquisition was performed as described in (Villasenor et al., 2020). In brief, samples (quadruplicates per condition) were cleaned up by C18 StageTips. Peptides were detected by data-dependent acquisition via mass spectrometry. Proteins were identified and quantified from the raw acquisition data as well as processed using MaxQuant (Cox and Mann, 2008). The mouse reference proteome (UniProtKB/Swiss-Prot) version 2018_12 combined with manually annotated contaminant proteins was searched with protein and peptide false discovery rate (FDR) values set to 1% and the match-between-runs algorithm was enabled. Statistical analysis was conducted using Proteus (bioRxiv 416511). LFQ intensity values were log2-transformed and outlier samples determined based on low peptide or protein counts. Subsequently, Proteus' *limma*-wrapper (Ritchie et al., 2015) was used to determine potential interactors in respective contrasts (bait vs. control of same genetic background) with an FDR of 0.05 as significance threshold. For gene ontology term analysis, differentially enriched proteins of bait conditions were combined and parsed to enrichr (Kuleshov et al., 2016).

7. Dynamics and Context-Dependent Roles of DNA Methylation

7.1. Published Review

Authors: Christina Ambrosi, Massimiliano Manzo, and Tuncay Baubec.

Published: Journal of Molecular Biology, 2017; doi: 10.1016/j.jmb.2017.02.008.
https://www.sciencedirect.com/science/article/abs/pii/S0022283617300839?via%3Dihub

Summary: In mammalian genomes, DNA methylation is the most common chemical modification of cytosines. It decorates regulatory elements of repressed genes, is known to control repetitive elements, and is also spread along actively transcribed gene bodies. Thus, with its dynamic nature it plays an important role in the establishment of cellular identity. Here, we reviewed how recent functional studies have helped to understand the differential distribution and effects of DNA methylation in different contexts during early development.

Author contribution: I participated in the conceptualisation and performed the majority of writing, figure and table preparation as well as revision of this review manuscript.

Number of Citations: 62

Review

Dynamics and Context-Dependent Roles of DNA Methylation

Christina Ambrosi[1,2], **Massimiliano Manzo**[1,2] **and Tuncay Baubec**[1]

1 - Department of Molecular Mechanisms of Disease, *University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
2 - Molecular Life Sciences PhD Program of the Life Sciences Zurich Graduate School, *University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

Correspondence to Tuncay Baubec:*at: Winterthurerstrasse 190, 8057 Zurich, Switzerland. tuncay.baubec@uzh.ch.*
http://dx.doi.org/10.1016/j.jmb.2017.02.008
Edited by Findlay Greg

Abstract

DNA methylation is one of the most extensively studied epigenetic marks. It is involved in transcriptional gene silencing and plays important roles during mammalian development. Its perturbation is often associated with human diseases. In mammalian genomes, DNA methylation is a prevalent modification that decorates the majority of cytosines. It is found at the promoters and enhancers of inactive genes, at repetitive elements, and within transcribed gene bodies. Its presence at promoters is dynamically linked to gene activity, suggesting that it could directly influence gene expression patterns and cellular identity. The genome-wide distribution and dynamic behaviour of this mark have been studied in great detail in a variety of tissues and cell lines, including early embryonic development and in embryonic stem cells. In combination with functional studies, these genome-wide maps of DNA methylation revealed interesting features of this mark and provided important insights into its dynamic nature and potential functional role in genome regulation. In this review, we discuss how these recent observations, in combination with insights obtained from biochemical and functional genetics studies, have expanded our current knowledge about the regulation and context-dependent roles of DNA methylation in mammalian genomes.

Introduction

DNA methylation is a prevalent chemical modification of cytosine bases (Fig. 1) and is found in many eukaryotic genomes [1,2]. It is associated with long-term transcriptional repression and is faithfully propagated during mitosis, suggesting a potential memory mechanism for transcriptional regulation [3,4]. In mammals, this mark is involved in the repression of gamete-specific genes during development [5] or entire chromosomes, for example, during mammalian X chromosome inactivation [6], and is essential for genomic imprinting by guiding allele-specific gene expression [7,8]. Furthermore, DNA methylation is required for the silencing of repetitive elements [9,10], thereby maintaining genome stability [11]. It is also considered to be important for cellular differentiation as well as identity and is of relevance to human health [12].

Although DNA methylation is one of the best mechanistically understood epigenetic modifications, the exact role of DNA methylation in regulating genome function and how this differs from gene to gene or genomic context need to be uncovered. To date, it still remains to be fully understood how specific DNA methylation patterns are precisely generated, maintained, read, or erased along the genome, in order to better understand its role as an epigenetic signal that regulates biological processes. The promising emergence of high-throughput sequencing has offered to explore this epigenetic mark at a genome-wide scale and at high resolution. DNA methylation measurements from various cell types and developmental stages, in combination with additional genome-wide datasets of chromatin modifications and transcriptional activity, have generated a great amount of knowledge and offer important insights into the dynamic regulation and function of this relevant modification. Building up on

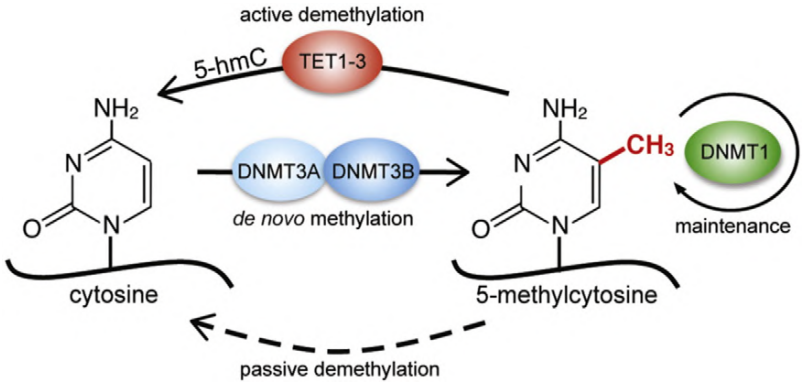


Fig. 1. Key players in the DNA methylation pathway. In eukaryotes, DNA methylation predominantly occurs at the fifth carbon atom of cytosine bases. Its deposition is catalysed by the *de novo* DNMTs DNMT3A and DNMT3B. Introduced methylation patterns are preserved by the maintenance DNMT DNMT1 during replication. Passive DNA demethylation is considered to be achieved across cell division in the absence of DNMT1 maintenance activity. Active removal includes the mammalian TET1–3 proteins that are capable of converting 5-methylcytosine to its oxidised derivative 5-hydroxymethylcytosine (5hmC) and further to 5-formylcytosine and 5-carboxylcytosine (not indicated here). These modifications are removed through DNA repair processes or are passively lost through replication. In addition, DNA repair processes have been involved in the direct removal of methylated cytosines.

this knowledge, this review focuses on general features of DNA methylation and how this modification is regulated and discusses the dynamics of genome-wide DNA methylation patterns in murine embryonic stem (mES) cells and during early mouse development.

DNA Methyltransferases (DNMTs), the Writers of DNA Methylation

In mammals, three conserved enzymes are responsible for the deposition of methyl mark to the fifth carbon atom of cytosine bases and are essential for normal development [13,14]. Therein, DNMT3A and DNMT3B catalyse *de novo* DNA methylation, whereas DNMT1 is the maintenance enzyme, which restores the fully methylated state of DNA after replication [15] (Fig. 1). In line with this role, DNMT1 prefers hemimethylated DNA as a substrate and is localised to the replication foci, ensuring the inheritance of methylation patterns [16–18]. Functional studies, however, have indicated that this distinction does not fully reflect their activities *in vivo*. For example, in mES cells lacking DNMT3A and DNMT3B, DNMT1 alone is not sufficient to faithfully maintain DNA methylation levels, indicating that the *de novo* enzymes are involved in either filling up gaps after DNA replication or are required to counteract active demethylation [19–21]. Recently, a new member of the DNMT family has been identified: DNMT3C. This novel DNMT evolved as a duplication of DNMT3B in rodents and is responsible for DNA methylation of retrotransposons during male germ cell development [22].

At a biochemical level, the structure of DNMTs and the mechanisms of establishment or maintenance of DNA methylation have been intensively studied over the last decades, bringing considerable insights into regulatory mechanisms that modulate their targeting to the genome or their enzymatic activity (reviewed in Refs. [23,24]) (Fig. 2a). All DNMTs share a similar, multidomain architecture, which can be distinguished into a variable N-terminal part containing regulatory domains and a C-terminal region harbouring the catalytic methyltransferase domain. The latter exhibits conserved amino acid motifs typical for DNMTs described in bacteria and is responsible for DNA binding and its chemical modification [15]. The N-terminal regulatory domains strongly differ between DNMT1 and DNMT3 proteins and could play an important role in specifying their activity along the genome. Most notably, the N-terminal part of DNMT1 contains domains that mediate anchoring to the replication fork [18,25] or specify the preference for hemimethylated substrates through autoinhibition upon binding to unmethylated DNA [26]. The N-terminal part of the DNMT3 methyltransferases contains two conserved domains. The ATRX-DNMT3-DNMT3L domain of DNMT3A, DNMT3B, and DNMT3L interacts with histone H3 tails and is blocked by the methylation of the lysine-4 residue (H3K4me3) [27–29], whereas the PWWP domain of DNMT3A and DNMT3B contains a conserved aromatic cage that recognises histone H3 lysine-36 trimethylation (H3K36me3) *in vitro* and *in vivo* [30–32] and is also required for the localisation to pericentromeric repeats [33]. In contrast, DNMT3C and DNMT3L lack the PWWP domain completely [22]. In addition to these conserved

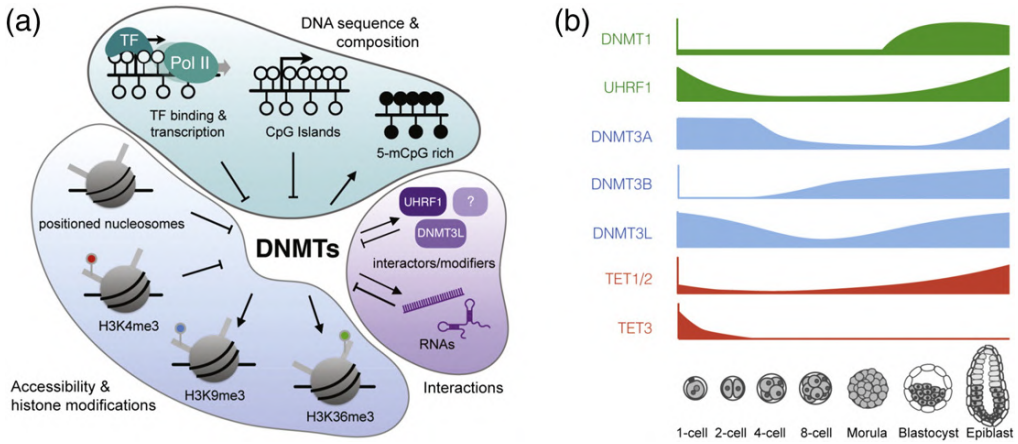


Fig. 2. DNMT targeting and activity. (a) Local and cell-type-specific patterns of DNA methylation can be achieved through the recruitment or repulsion of DNMTs to specific genomic sites. Interactions with other proteins such as UHRF1 or DNMT3L or various RNA molecules can promote or prevent DNMT targeting to genomic sites. Post-translational modifications of histone H3 tails can also alter DNMT targeting preferences *in vitro* and *in vivo,* either directly or indirectly. These mediate local DNMT activity through increased repulsion (H3K4me3) or retention (H3K9me3 and H3K36me3) to sites enriched for these marks. Nucleosome positioning (accessibility), bound TFs, or engaged RNA polymerases II (Pol II) can interfere with DNMT recruitment and thus decrease local DNMT activity. Densely methylated and unmethylated CpG-rich regions (CpG islands) in the genome either promote or block DNMT recruitment and activity. (b) The DNA methylation-modifying enzymes DNMTs and TETs and their major accessory proteins, DNMT3L and UHRF1, display distinct activity patterns throughout early development.

domains, the most N-terminal part of DNMT3A and DNMT3B is highly variable and is involved in interactions with DNA and nucleosomes [34,35]. The shorter isoform of DNMT3A that is predominantly expressed in embryonic stem (ES) cells, DNMT3A2, lacks this variable part [36]. Understanding whether and how these domains work together to specify genomic localisation and activity of DNMTs is still a matter of current investigations that should bring valuable insights into the mechanisms that establish cell-type-specific DNA methylation patterns. Several biochemistry-based studies have started to shed light on how these domains could integrate chromatin information to guide DNMTs to specific genomic sites (reviewed in Ref. [24]); however, additional studies are required that incorporate this knowledge from a genomics perspective.

In addition, post-translational modifications [37,38] and interactions with other proteins such as ubiquitin-like with PHD and ring finger domains 1 (UHRF1) [39], the ATP-dependent chromatin remodeller lymphoid-specific helicase (LSH) [40], the inactive DNMT3 family member, DNMT3L [41], or various RNA molecules [42–44] have been implicated in regulating DNMT-specific targeting preferences and activities *in vivo* and *in vitro* (Fig. 2a). Most notably, interactions between DNMT1 and UHRF1 are essential for the maintenance of DNA methylation during replication [39,45]. Immediately after DNA replication, the newly synthesised DNA strand lacks methylation. UHRF1 binds this hemimethylated DNA through its SRA domain and recruits DNMT1 to the replication fork [46,47]. Furthermore, UHRF1 has been found to interact with histone H3 lysine 9 (H3K9) methylation via its tandem tudor domain, linking H3K9 methylation to DNA methylation maintenance [48,49]. Interactions between *de novo* DNMTs with their catalytic inactive family member DNMT3L increase their enzymatic activity and targeting specificity [50–52]. Most remarkably, DNMT3L plays an essential role in regulating DNA methylation in gametes, where it guides DNMT3A to establish genomic imprints and grants genome stability through the methylation of transposable elements [11,53–55]. Despite its relevance during gamete maturation, DNMT3L is dispensable during early mouse development [56].

The catalytically active DNMT family members play an essential role during mammalian development, and *Dnmt* knockout mice are lethal (Table 1). However, the basis for this lethality still remains to be identified. Interestingly, the phenotypes of individual *Dnmt* gene ablations differ significantly. Whereas gene knockouts of *Dnmt1* and *Dnmt3b* lead to embryonic lethality, the knockout of *Dnmt3a* is embryonic viable and only dies approximately 4 weeks after birth [13,14,57].

The observed phenotypes match the expression patterns and essential roles of DNMTs in the developing mouse embryo (Fig. 2b). ES cells express different isoforms of DNMT3 proteins, predominantly the isoforms DNMT3A2 and DNMT3B1 [14,20,58]. During later development, DNMT3A is transcribed from an alternative promoter

Table 1. Relevance of DNA methylation writers and erasers for mouse development

Genes	Knockout studies in mice
Dnmt1	Embryonically lethal at E9.5, distorted neural tube, lacks somites [13,57]
Dnmt3a	Normal development until birth, postnatally (4 weeks) rudimentary and fatal [14]
Dnmt3b	Embryonically lethal at E 9.5–10.5, growth impairment, neural tube defects [14]
TET1	Vital, reduced postnatal body size, mild developmental delay [84]
TET 2	Normal development; at 20 weeks, defects in blood differentiation and elevation of extramedullary haematopoiesis [83]
TET3	Reduced litter size, arrest around E11.5, perinatal death [85]

DNA methylation is known to be an essential and highly dynamic feature throughout development and differentiation. Knockouts of individual enzymes involved in the setting and removing of DNA methylation display a wide range of phenotypes, ranging from early lethality to reduced body size. This table summarises the phenotypes obtained from mouse knockout studies.

and produces the longer isoform DNMT3A1. This becomes the predominant isoform of DNMT3A in differentiated tissues [36]. Additionally, catalytically inactive isoforms of DNMT3B are expressed at various stages of development, which could potentially influence methylation patterns through interactions with catalytically active DNMT3A or DNMT3B [52,59]. Although DNMT3A apparently plays a minor role in the contribution of DNA methylation to the developing embryo, DNMT3A, together with DNMT3L, is essential for the establishment of DNA methylation patterns in gametes, including maternal and paternal genomic imprints [8,53,54,60]. Similarly, DNMT3C regulates young retrotransposons in the male germline and is required for mouse fertility [22]. DNMT3B, on the other hand, plays a more important role in methylating target promoter regions during embryonic development and in setting DNA methylation genome-wide, including repetitive elements during implantation [5,14].

Erasure of DNA Methylation Patterns

Since DNA methylation is highly dynamic and varies in a global and local context, DNA methylation removal is considered as an integral part of the epigenetic regulatory network. DNA methylation can be removed through several mechanisms (reviewed in Ref. [61]). Passive DNA demethylation is achieved via DNA replication in the absence of DNMT1 activity (Fig. 1). During replication, hemimethylated CpG intermediates are produced, and DNMT1 copies the methylation state of the mother strand onto the newly synthesised strand. In absence of maintenance DNA methylation activity, CpG methylation gets diluted during replication and can be completely absent from a locus after two cell divisions.

Alternatively, several processes have been proposed to be involved in the active removal of methylation in a site-specific or global manner, such as DNA repair [62,63] or oxidation. The mammalian Ten-Eleven Translocation (TET) proteins are capable of converting 5-methylcytosine to its oxidised derivative 5-hydroxymethylcytosine (5hmC) and further to 5-formylcytosine and 5-carboxylcytosine [64–66]. These oxidised products of 5mC can be actively and passively removed by DNA repair and through cell division, respectively [67,68]. Three *Tet* family members exist in mammalian genomes, *Tet1–3* [69]. All three TET proteins are involved in various biological processes, ranging from embryonic development to gene regulation and cancer (reviewed in more detail in Ref. [70]). Based on their catalytic activity, TET proteins contribute to the regulation of DNA methylation patterns, and this requires precise recruitment to genomic elements. TET1 and TET3 contain N-terminal CXXC domains, and TET2 interacts with the CXXC domain protein IDAX [71,72]. CXXC domains preferentially bind to unmethylated CpGs [73], providing a potential mechanism of TET recruitment to unmethylated CpG-rich promoters [72,74–76]. However, how TETs are recruited to the genome and to what extent CXXC domains contribute to their specificity are still a matter of debate. Genome-wide studies revealed the preference of TET proteins to bind to CpG islands (CGIs) [77,78]. However, TETs also prefer to bind to accessible and low methylated sites in the genome with lower CpG densities [79], and the abundance of 5hmC at enhancer elements and methylated gene bodies [80,81] suggests that additional factors could promote site-specific recruitment of these proteins.

Individual knockouts of TET proteins have less severe outcomes on development compared to knockouts of DNMTs (Table 1). *Tet1* and *Tet2* are expressed during early development in the inner cell mass (ICM) and during implantation, whereas *Tet3* is expressed in the mouse oocyte and zygote until the 2-cell stage and has been suggested to contribute to the demethylation of both parental genomes [82] (Fig. 2b). Knockout of *Tet1* and *Tet2* in mice generated from crosses between heterozygotes led to viable progeny, whereas loss of TET1 results in reduced litter size, and loss of TET2 promotes leukaemias [83–84]. *Tet3* knockouts generated by conditional deletion in the germline lead to perinatal death [85] (Table 1). Follow-up studies using *Tet1–2* double-knockouts indicated that their removal is compatible with development [86], while triple--knockouts of *Tet1–3* genes demonstrated defects in chimera contribution and proper differentiation

[87]. In mouse ES cells, *Tet1,2*-double and *Tet1,2,3*-triple knockouts resulted in the global depletion of 5hmC and increased DNA methylation at enhancers, without apparent implications for self-renewal [88,89].

Conservation, Global Distribution, and Sequence Context of DNA Methylation

In contrast to plants and invertebrates, where genome-wide DNA methylation patterns follow a "mosaic" distribution, vertebrate methylation occurs throughout the entire genome [1,2]. Methylation is present mainly in the context of CpG dinucleotides, although non-CpG methylation is evident in mammalian tissues, especially in ES and neuronal cells [90,91]. In mammalian genomes, 60–80% of CpG sites are methylated, and roughly 10% occur in CGIs [92]. CGIs are marked by a higher density of CpG dinucleotides than expected by chance, and the majority of CGIs are associated with active transcriptional units and, to a great degree, are resistant to methylation [93–95]. Unmethylated CGIs are considered to prohibit DNA methylation through various processes, including engaged RNA polymerase II [96], transcription factor (TF) binding [97], and H3K4me3 [27,29]. Furthermore, binding of proteins involved in H3K4me3 deposition (CFP1/SET1, MLL1,2) or removal of DNA methylation by TETs could provide a positive feedback loop that stabilises the unmethylated state [76,98,99]. Approximately 60% of transcriptional start sites coincide with CGIs. These include promoters of housekeeping genes, which are active across various cell types [100]; however, a small fraction of CGI promoters that control imprinted and tissue-specific genes are methylated [7,92].

Spontaneous and frequent deamination of methylated cytosines to thymine is the major cause of global CpG depletion during the evolution of mammalian genomes [101,102]. The retention of a hypomethylated state at promoters can generally be seen as a protective mechanism that prevents the erosion of CpGs by methylation. Overall, the distribution of CpGs in mammalian genomes is highly skewed and widely shaped by methylcytosine deamination, selection pressure (e.g., CpGs in codons), and active protection from methylation through DNA-sequence-dependent mechanisms [96,102–105]. This skewed CpG distribution can influence the regulatory potential of DNA methylation through local variation in methylation density. For instance, methylated CpG-rich promoters are more likely to be regulated by DNA methylation and are more frequently bound by methyl-CpG-binding domain (MBD) proteins [95,103,104].

A Genomics View on DNA Methylation Dynamics

The last decade has generated a great amount of information related to the distribution of methylated cytosines in mammalian genomes. Fuelled by increasing sequencing depth and decreasing sequencing costs, the first comprehensive mammalian whole-genome bisulphite sequencing maps were obtained from human ES, fibroblasts, and sperm and haematopoietic cells [91,105,106], followed by mES and neuronal progenitor cells [79] and an entire panel of developmental transitions in mouse and human [5,107,108]. These datasets did not only measure the frequency and distribution of individual methylated cytosines in the mammalian genome but also greatly contributed towards understanding how the deposition of DNA methylation is regulated and how it could influence gene activity. Most importantly, comparison of DNA methylation profiles to other genome-wide datasets extracted from the same tissues and cell types provided valuable insights about the interplay among DNA methylation, DNA sequence, transcriptional activity, and other histone modifications. Many of these comparisons confirmed previous findings at a genome-wide scale and unprecedented resolution, generating useful resources for the community, but also led to new discoveries important for our current understanding of DNA methylation and its function.

Profiles of DNA methylation obtained from early developmental transitions or gamete maturation revealed detailed insights into the dynamic nature of DNA methylation during this important stage of development, where global DNA methylation reprogramming takes place (Fig. 3a). In addition, studies using mES cells as cell culture models that allow transitions between naïve and primed states or terminal differentiation were used to dissect numerous mechanisms that regulate DNA methylation deposition and erasure at a genome-wide scale. Below, we will mainly discuss how these recent genome-wide studies have contributed to our current understanding about the regulation and function of DNA methylation during these early developmental stages and in ES cells.

Establishment of Gamete-Specific DNA Methylation Patterns

The first wave of genome-wide DNA methylation resetting takes place during primordial germ cell (PGC) development. During PGC migration into the genital ridge, DNA methylation marks are removed from the entire genome, with the exception of a subset of repetitive elements that retain elevated

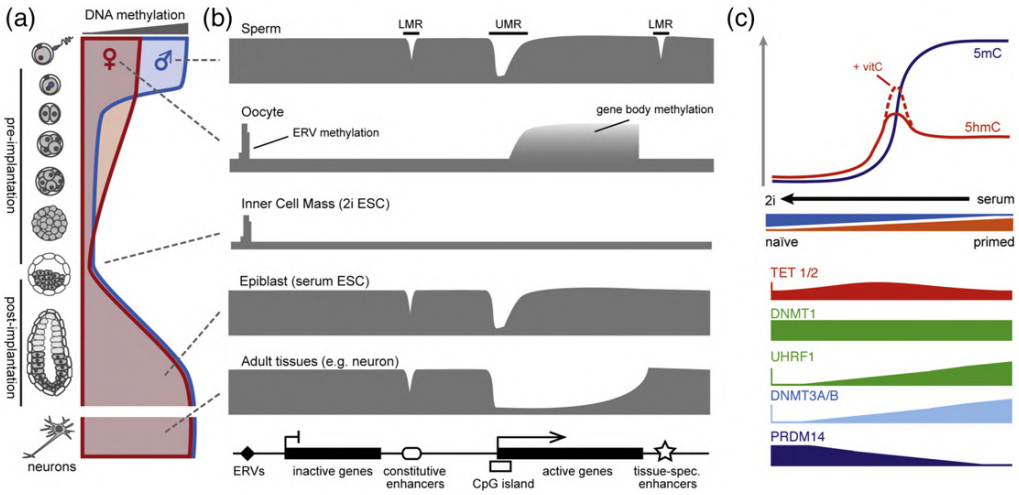


Fig. 3. DNA methylation dynamics during early development. (a) A classical view on DNA methylation dynamics during early mouse development indicating results obtained from bulk measurements of maternal (red) and paternal (blue) methylation. This indicates the major wave of demethylation in preimplantation embryos that starts after zygote formation and culminates at the blastocyst stage. DNA methylation is reestablished during implantation. (b) Representative view of DNA methylation landscapes during early mouse developmental stages and in neurons as a comparison. Shown are DNA methylation patterns at active and inactive genes, constitutive and tissue-specific enhancers, and endogenous retroviruses (ERK) as a representative repetitive element that maintains DNA methylation during mouse preimplantation development. Specific DNA methylation patterns that occur globally or during specific stages and cell types are indicated. These include low methylated regions (LMRs) that coincide with active distal regulatory regions, unmethylated regions (UMR) coinciding with active and CGI promoters, and methylated and unmethylated gene bodies of actively transcribed genes. (c) mES cells cultivated in leukaemia-inhibitory factor and serum-containing media exhibit high levels of DNA methylation (5mC; >80%) and are more "primed" towards a differentiated state. In the presence of Erk1/2 and GSK3beta inhibitors (2i) ES cells can be maintained in a homogenous and naïve state, closely resembling the ICM. Adaptation to 2i results in rapid global reduction of DNA methylation. Global and rapid demethylation is achieved through decreased DNA methylation maintenance upon the degradation of UHRF1 and the transcriptional downregulation of the *de novo* DNMTs DNMT3A and DNMT3B by PRDM14. Global demethylation is enhanced by the activity of TET dioxygenases, forming 5-hydroxymethylcytosine (5hmC) intermediates, which is further boosted in presence of vitamin C.

DNA methylation levels throughout the entire duration of PGC development [62,109–112]. After this genome-wide erasure of DNA methylation, the genome becomes remethylated. Remethylation takes place in different cellular contexts and at different developmental timepoints for female and male gametes in mice. Therefore, rates and generated patterns of DNA methylation differ strongly between sperm and oocyte [60,113,114] (Fig. 3a and b). Remethylation in the male gametes initiates in G1-arrested prospermatogonia and is largely established around birth, resulting in the majority of CpGs being remethylated in mature sperm, with the exception of those overlapping CGIs, active promoters, and active enhancers [105,115,116] (Fig. 3b). The majority of CpGs are already fully methylated in postnatal prospermatogonia, with little additional increase in methylation until full maturation [116]. However, during murine germ cell development, numerous stage-specific, differentially methylated regions were identified that overlap with distal regulatory elements and contain binding sites for stage-specific TFs. This finding is, however, in contrast to an earlier study reporting little changes in DNA methylation during spermatogonial stem cell differentiation [115].

In mice, DNA methylation in female gametes initiates around birth during follicular growth of meiotically arrested oocytes [117]. Since no cell division occurs during this global remethylation, the DNA methylation patterns in the mature oocyte largely reflect *de novo* DNA methylation activity of DNMT3A, together with DNMT3L [8,53,60]. This provides important insights about the regulatory mechanisms involved in targeting *de novo* methylation to the genome during oocyte maturation. Surprisingly, methylation in mature mouse oocytes is mosaic and restricted to defined genomic regions overlapping with transcribed genes and repetitive elements, while the rest of the genome remains hypomethylated, resulting in 40% CpG methylation overall [60,114,118] (Fig. 3b). Several studies in mice suggest that histone modifications play an important role in guiding *de novo* methylation in oocytes. Increased methylation during oocyte maturation coincides with actively transcribed intragenic CGIs, suggesting a relationship between transcription and *de novo* methylation

[60,114]. Comparative analysis of co-transcriptionally deposited H3K36me3 with DNA methylation in growing murine oocytes further supports a direct role of transcriptional elongation and H3K36me3 in guiding *de novo* DNA methylation [119]. However, a direct role for H3K36me3 in guiding methylation patterns in oocytes remains to be shown. Regions occupied by H3K4 di- and trimethylation are protected from *de novo* methylation in oocytes, supporting the antagonism between H3K4 methylation and DNMT3A/DNMT3L mediated by their ATRX-DNMT3-DNMT3L domains [27,29]. Recent genome-wide studies further identified broad H3K4me3 domains in mature oocytes, which differ in size and genomic localisation compared to the canonical H3K4me3 peaks identified in mouse ES and somatic cells. These domains overlap with hypomethylated regions in oocytes, suggesting that this unusual H3K4me3 distribution could be involved in the large-scale genome hypomethylation observed in oocytes through preventing *de novo* methylation [120,121]. However, since the establishment of DNA methylation and non-canonical H3K4me3 domains seems to be temporary correlated during oocyte growth, additional functional studies are required to uncover the causal relationship between these marks. Here, it is relevant to mention that while all H3K4me3 domains coincide with hypomethylated regions, only 60% of hypomethylated regions reside within the non-canonical H3K4me3 domains [120,121]. Therefore, additional mechanisms could exist that protect the remaining regions from *de novo* methylation.

Resetting DNA Methylation Patterns during Early Embryogenesis

The pervasive, differential methylation established in the mature gametes [60,114] is largely resolved by the genome-wide removal of DNA methylation that initiates after fertilisation (Fig. 3a). With a few exceptions, the removal of DNA methylation seems to occur at equal rates over the entire genome, independent of the functional element annotation or CpG density [122,123]. However, large differences exist between the oocyte and sperm demethylation rates. Early cytological work in mice indicated differential DNA methylation reprogramming dynamics of the parental genomes [124–126]. The paternal genome displays active and almost complete demethylation after fertilisation and prior to replication, whereas the maternal genome appears to become demethylated passively during cell divisions, reaching its minimum around the blastocyst stage before implantation (Fig. 3a). How DNA methylation is differentially removed from the parental genomes is not completely elucidated, but similar to the demethylation processes during PGC development, a combination of various pathways could facilitate genome-wide removal. These include TET-mediated oxidation [85,127,128], DNA repair pathways [63,129], and suppression of DNA methylation [130,131]. Increased 5hmC localisation to the paternal genome after zygote formation indicates that TET-mediated oxidation could be accounted for the active removal of DNA methylation [85,127,128]. However, these studies were based on immunofluorescence detection, and more sensitive measurements using mass spectrometry revealed that the rate of paternal demethylation is not substantially affected in absence of TET3-mediated oxidation [132]. Surprisingly, the same study identified that the occurrence of 5hmC on the paternal genome after fertilisation depends on DNMT1 and DNMT3A, suggesting that TET3-mediated oxidation is required to prevent DNA methylation during this early stage. Nevertheless, the exact mechanism that leads to the active removal of methylation from the paternal genome is still a matter of debate.

Due to this difference in demethylation rates, the majority of DNA methylation patterns measured through genomics assays in the early mouse embryo are largely contributed by the inherited oocyte methylation, although these patterns are diluted with ongoing cleavage stage until full removal [114,122,123] (Fig. 3a and b). While the bulk of DNA methylation is removed during preimplantation, some genomic sites remain protected from demethylation and carry the methylation state established in the gametes into the late embryo [123]. These include mostly endogenous retroviruses (ERVs) of the intracisternal A-particle and differentially methylated imprinting control regions [5,110,133]. Maintenance of DNA methylation at these regions requires DNMT1 activity [14,57]. A maternally inherited, shorter isoform of DNMT1, DNMT1o, prevents the loss of methylation at intracisternal A-particles during cleavage, while the embryonic full-length DNMT1 contributes to methylation maintenance during later stages [14,57,134].

Following this ground state of DNA methylation at the blastocyst stage [135], *de novo* DNMTs are upregulated, and DNA methylation is reestablished genome-wide during the transition to the epiblast stage [5,14,136,137] (Fig. 3). Absence of DNA methyltranferases during this stage results in limited remethylation and postimplantation lethality, whereas *de novo* methyltransferases—mainly DNMT3B—carry out the establishment of methylation, and DNMT1 is propagating the established patterns during cell division [5,14,57]. As already pointed out by Okano and colleagues [14], the *de novo* methyltransferases display different expression patterns, activities, and targeting preferences during postimplantation development. While DNMT3B seems to be largely responsible for the

majority of *de novo* methylation events during implantation, DNMT3A is more relevant during later differentiation [5,8,14,138,139]. During implantation in mice, several promoters become *de novo* methylated, including pluripotency gene promoters, promoters of gamete-specific genes, and numerous promoters associated with tissue-specific genes that are active during later development [5,140]. DNA methylation at CpG-dense, gamete-specific gene promoters is maintained through the later stages of development, and interference with DNA methylation during gastrulation or in various other cell types results in their transcriptional reactivation [5,141,142]. This indicates that the high density of methylation at these genes could play an important role in preventing aberrant transcription [95,103]. In addition, the promoters of pluripotency genes, including Oct4 and Nanog, become *de novo* methylated in a stage-dependent manner, suggesting a role for DNA methylation in coordinating transcriptional programs during exit from pluripotency [5,140,143].

DNA Methylation and Pluripotency

ES cells are derived from the ICM, can be cultivated for unlimited passages, have the unique property to differentiate to various cell types and contribute to all germ layers [144]. This self-renewal capacity and pluripotency make them a valuable model for studying the role of epigenetic modifications during early development. mES cells cultivated in leukaemia-inhibitory factor and serum-containing media display high levels of DNA methylation (>80%) [79,91]. This state of full-genome methylation is in contrast to the hypomethylated genome of the ICM (Fig. 3). Cultivation of mES cells in a defined, serum-free medium and in the presence of Erk1/2 and GSK3beta inhibitors (2i) allows mES cells to maintain a state that, based on transcription and DNA methylation, closely resembles the ICM [135,145–149] (Fig. 3c). It is widely accepted that mES cells cultivated in presence of serum are more "primed" towards a differentiated state and constantly cycle between self-renewal and differentiation-prone states, while 2i-grown mES cells resemble a homogenous, naïve state [145,146]. Recent single-cell gene expression analysis identified that serum-grown mES cells display increased variation of pluripotency gene expression, confirming previous findings [150], whereas this variation is not observed in 2i cells [151].

This heterogeneity in the presence of serum also affects DNA methylation. Serum-grown mES cells show reduced clonal transmission of DNA methylation patterns, which could be attributed to increased turnover rates of DNA methylation at regulatory elements [152,153]. This creates increased heterogeneity in DNA methylation, which can be observed from bulk or single-cell whole-genome bisulphite sequencing data [79,154]. These sites of heterogeneity often coincide with distal regulatory elements enriched for pluripotency and differentiation factors. Despite these correlations, it remains to be clarified if heterogeneity of DNA methylation at these sites could play any role in regulating the dynamic activity of pluripotency and differentiation factors. While global interference with DNA methylation has been reported to reduce the gene expression variability of pluripotency genes [155], additional studies are required to fully substantiate a role for methylation in regulating expression dynamics of pluripotency genes. Importantly, complete depletion of DNA methylation in the absence of DNMT1, DNMT3A, and DNMT3B (Dnmt-triple-KO) does not alter the morphological features or self-renewal properties of mouse ES cells, suggesting that methylation is not generally required for the maintenance of an ES cell state [141,156]. However, comprehensive single-cell RNA sequencing experiments are required to identify if pluripotency genes in mouse Dnmt-triple-KO ES cells recapitulate the transcriptional heterogeneity of serum mES cells or are more similar to mES cells grown in 2i conditions.

The transition between naïve and primed states of mES cells can be easily controlled in cell culture, providing a useful tool to study the mechanism of DNA methylation deposition and removal (Fig. 3c). Adaptation of serum-grown ES cells to 2i or *vice versa* results in rapid global reduction or reestablishment of DNA methylation after 10 days, respectively [148]. How this rapid removal of cytosine methylation is facilitated was the focus of numerous recent studies (Fig. 3c). Earlier studies suggested that the transcriptional downregulation of *Dnmt3a* and *Dnmt3b* by PRDM14 upon cultivation in 2i medium could be responsible for the global demethylation [135,148,149,157]. However, this hypothesis was contradicting reports using *Dnmt3a* and *Dnmt3b* double-KO mES cells, where global reduction of DNA methylation required extensive passaging [20,21]. A potential explanation to this discrepancy involved the enhanced TET-mediated removal of DNA methylation during cultivation in 2i [158]. Interestingly, addition of vitamin C further accelerated DNA methylation removal through boosting the activity of TET enzymes and temporally increasing 5hmC [159]. Surprisingly, recent findings indicate that TET enzymes do not substantially contribute to the global removal of methylation during the transition from serum to 2i, but rather that the decreased DNA methylation maintenance through downregulation of UHRF1 is responsible for the global reduction [160], in line with the role of UHRF1 in the recruitment of DNMT1 to hemimethylated DNA [39,45].

DNA Methylation during Exit from Pluripotency and Differentiation

The ability of ES cells to differentiate to various cell types provides a useful model to study the regulation of DNA methylation dynamics during lineage commitment and terminal differentiation, as well as the requirement of DNA methylation for exit from pluripotency and differentiation. Earlier studies comparing DNA methylation profiles of mouse ES cells to differentiated cell types, and during *in vitro* differentiation, detected dynamic changes of this mark at numerous promoters and enhancers [79,107,143,161]. Similar to *de novo* methylation events during gastrulation, promoters and enhancers of pluripotency genes become *de novo* methylated, suggesting that methylation could contribute to exit from pluripotency or prevent dedifferentiation. This establishes the model that DNA methylation, similar to other epigenetic modifications, could serve as a barrier to cellular reprogramming and could contribute to the low efficiency of induced pluripotent stem cell generation [162]. Along these lines, successfully reprogrammed induced pluripotent stem cell cells are characterised by hypomethylated pluripotency gene promoters, and interference with DNA methylation has been reported to improve reprogramming efficiency [163,164]. This suggests that DNA methylation at promoters and enhancers of pluripotency genes could block the accessibility of the pluripotency factors OCT4, KLF4, SOX2, and cMYC to the DNA and therefore interfere with the establishment of a pluripotency gene regulatory network. The requirement of *de novo* DNA methylation at promoters of pluripotency genes during exit from pluripotency is, however, less clear. Strikingly, early passage mouse ES cells lacking *Dnmt3a* and *Dnmt3b* are able to terminally differentiate *in vitro*, suggesting that *de novo* methylation of pluripotency genes is dispensable for exit from pluripotency and further commitment [21]. In contrast, mouse ES cells lacking DNMT1 fail to differentiate, suggesting that the maintenance of methylation patterns, rather than *de novo* establishment, is required for differentiation [21,165]. A similar failure to differentiate is also observed for late-passage *Dnmt3a* and *Dnmt3b* double-knockout mES cells, which progressively lose methylation due to their involvement in “backing up” maintenance methylation [20,21]. Why the maintenance of methylation is more important than *de novo* methylation during differentiation is not fully understood, but it could be required to maintain transcriptional repression at either gamete-specific gene promoters or repetitive elements

Strikingly, mES cells retain their self-renewing capacity in the absence of DNA methylation in *Dnmt*-TKO cells [141,156]. While several imprinted and gamete-specific genes become derepressed in *Dnmt*-TKO mES cells, repetitive elements remain largely inactive [141,166]. This is in contrast to somatic and gametic cells, where the lack of DNA methylation leads to the upregulation of ERVs, resulting in genome instability [10,11]. DNA methylation is therefore dispensable for the maintenance of genome integrity in mouse ES cells, but upon exit from pluripotency, the absence of DNA methylation could trigger the reactivation of repetitive elements, resulting in interference with transcriptional programs and genome instability [10,166]. Genome stability in mES cells could be due to a dual backup scheme that involves H3K9me3. Removal of H3K9me3 in mES cells through the ablation of ESET, the responsible histone methyltransferase, or KAP1, a factor involved in the recruitment of various repressive factors to ERVs, leads to drastic upregulation of ERV-specific transcripts and rapid cell death, suggesting that H3K9me3 is the major protective mechanism in mES cells [166–169]. In line with a role for histone modifications in the maintenance of genome integrity, recent studies using the transition of mES cells from serum to 2i in the presence of vitamin C identified a transient upregulation of numerous repetitive elements upon rapid removal of DNA methylation, followed by a reconfiguration to H3K9me3- and H3K27me3-based repressive mechanisms [170]. This remarkable adaptation potential of mES cells in the absence of DNA methylation could reflect the epigenetic plasticity that takes place during this unique window of early mouse development. Further epigenetic perturbation experiments in ES cells are expected to reveal additional mechanistic insights into the regulation and crosstalk of DNA methylation with other histone modifications.

DNA Methylation - Instructive or Ensuing?

Generally, the regulation of gene expression is encoded in *cis* and is mediated by spatiotemporal interactions of TFs with their recognition sites within proximal and distal regulatory elements [171]. Nevertheless, gene expression can also be directed and modulated by epigenetic modifications through “overriding” genetically encoded programs in order to configure developmental decisions [172]. DNA and histone modifications correlate with transcriptional processes and are generally viewed to be involved in gene regulation. However, evidence for the direct contribution of epigenetic marks is still sparse. In the case of DNA methylation, presence of this mark at promoters is associated with the lack of gene activity, which could be mediated either through attraction or repulsion of regulatory factors

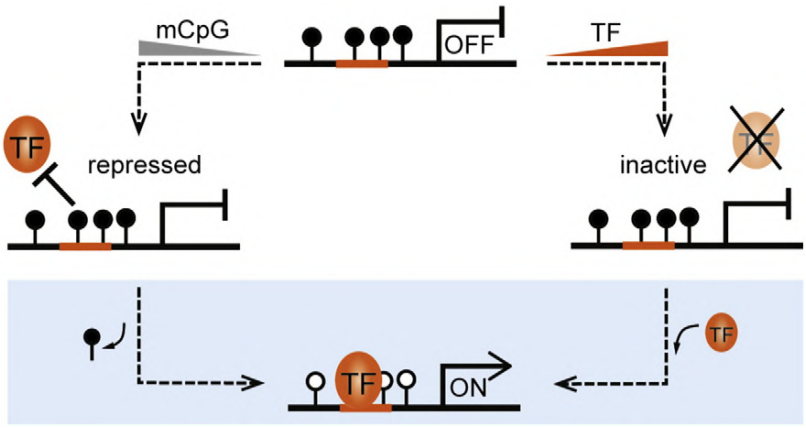


Fig. 4. DNA methylation, instructive or ensuing? DNA methylation at gene promoters is associated with transcriptional repression. However, whether the presence of methylation is the cause or consequence of gene inactivity is not always clear and could be mainly mediated by local context. For example, increased CpG density at methylated promoters is often found at gamete-specific genes that are reactivated in the absence of methylation. In this case, the density of methylated CpGs could directly or indirectly block TF binding and therefore lead to gene repression (left route). On the other hand, the presence of TF binding sites and the affinity and abundance of TFs to their recognition sites have been shown to influence DNA methylation at CpG-poor promoters and enhancers. In this case, the absence of an activating signal (i.e., TF activity) would render a gene inactive, and its promoter will become methylated by default (right route). Changes in DNA methylation patterns during biological processes, such as responses to stimuli or cellular differentiation, are often associated with changes in gene expression (blue box). For the repressed gene, removal of DNA methylation is required for TF binding and gene activation. In case of the inactive gene, changes in TF activity would result in enhanced TF binding to its recognition sequence and removal of DNA methylation through active or passive processes.

[104]. This correlative observation is true for the entire genome, but it does not imply that every methylated promoter is indeed repressed by DNA methylation (Fig. 4). For example, knockout of all DNMTs, and therefore, the complete removal of DNA methylation in mouse ES cells, results in the reactivation of only a small fraction of methylated promoters [141,156,166]. On the other hand, mES cells lacking DNA methylation are unable to differentiate [21,173]. These findings suggest that DNA methylation-mediated regulation could be largely dependent on genomic context, such as variation in CpGs or TF sites at promoters, and/or cellular context, such as abundance of transcriptional regulators at different developmental stages.

DNA methylation at promoter sites is considered to lock genes in a stable silent state. Several studies have shown that methylated CpGs influence TF binding at their recognition site [174–177]. However, repulsion by DNA methylation seems to vary from TF to TF, and more comprehensive *in vivo* studies are required to uncover if methylated CpGs can generally block TF accessibility. Numerous TF recognition sites do not contain CpGs in their recognition sequence; therefore, direct repulsion is not expected to be a universal mechanism. It has to be further noted that this “traditional” antagonism between TF binding and DNA methylation has been blurred by recent reports that identified *in vitro* affinities of numerous TF to methylated DNA [178–180]. However, a direct requirement for DNA methylation in the recruitment of these TFs *in vivo* awaits confirmation.

DNA methylation can attract MBD proteins in a sequence-independent fashion [181,182]. Binding of MBD proteins could directly prevent TFs from binding through steric interference or through recruitment of chromatin remodelers and repressors [183–185]. However, this requires a high local density of methylated CpGs, which might also explain the observed repressive activity of methylation at CpG-rich promoters [95,103,186]. Despite the preference of MBD proteins for methyl-CpG-rich regions in the genome [184,186], it remains to be tested if MBD-interacting chromatin regulators indeed colocalise to the same genomic regions. In general, the role of MBD proteins in mediating methylation readout and gene regulation remains to be explored in more detail through genetic and genomic approaches. So far, studies of individual and combined genetic deletions of MBD family members did not show an expected reactivation of methylated promoters and left many questions about their role and redundancy in embryonic development unanswered [187,188].

Synthetic tools to directly manipulate DNA methylation at specific loci have already been described in 1997 [189], and more recently, with the developments of CRISPR/Cas9, site-specific targeting of DNA methylation writers and erasers can be easily achieved. For example, synthetic TFs engineered

from enzymatically inactive Cas9 fused to DNMT3A or TET1 could be utilised to target DNA methylation or demethylation to specific genes and regulate their transcription [190–192]. Furthermore, a similar system could be used to assess the effects of combinatorial recruitment of epigenetic regulators on genomic sites [193]. In this study, the authors could further show that DNA methylation-mediated silencing was maintained upon the removal of the synthetic TF. However, the extent of downregulation differed between different targeted genes, indicating a context-dependent effect of transcriptional silencing [193]. Such synthetic approaches provide an excellent opportunity to test the direct contribution of DNA methylation to gene regulation in a locus-specific manner. However, more comprehensive studies that query multiple sites in the genome are needed in order to understand the context-dependent influence of DNA methylation on transcription.

Protein–DNA Interactions Influence DNA Methylation Patterns

As noted above, how transcription factors are influenced by DNA methylation is not fully understood and can vary from factor to factor. On the other hand, single-locus and genome-wide studies have indicated that TF binding can protect from DNA methylation at proximal and distal regulatory regions in a tissue-specific manner [79,91,194,195]. Human and mouse ES cell data identified that binding sites of pluripotency factors show reduced DNA methylation [79,91]. Unbiased analysis over the entire genome identified the occurrence of low methylated regions that overlap with distal regulatory sites [79] (Fig. 3b). These regions display average DNA methylation values of 15–50%, increased conservation scores, DNaseI hypersensitivity, and TF binding sites. These observations gave rise to several important questions. Does this reduction in methylation at TF binding sites reflect a requirement for the removal of methylation prior to binding or a footprint of TF binding? Is removal mediated through the recruitment of active demethylation mechanisms to TF binding sites or through a consequence of decreased accessibility of the methylation machinery to these sites? 5hmC levels are drastically enriched at distal regulatory sites [79,196], implicating TET proteins in facilitating the removal of methylation upon or prior to TF binding. The role of TET proteins in promoting 5hmC at mES cell enhancers has been confirmed through knockout studies [88,89], but the requirement for this oxidation step in mediating TF binding remains to be tested. Nevertheless, the presence of 5hmC at TF binding sites suggests that constant methylation and demethylation take place at these regions, keeping DNA modifications highly dynamic. Based on these observations, targeted removal of DNA methylation, as described above, could indeed predefine which of the numerous TF motifs that usually occur throughout the entire genome should be occupied and which should not. Surprisingly, in mouse ES cells lacking DNA methylation, chromatin accessibility measured by DNaseI hypersensitivity is not drastically changed, suggesting that the majority of TFs expressed in mES cells are not sensitive to DNA methylation and/or that demethylation is not a prerequisite for their initial binding to methylated sites [177,197]. However, notable exceptions exist, as these studies identified that the binding of NRF1, GABPA, and MYCN to various genomic sites is blocked by the presence of DNA methylation in their binding motifs. These sites are only accessible in a methylation-free state or upon the binding of pioneering TF in the vicinity [177], again exemplifying how local context influences the role of DNA methylation in guiding TF binding to the genome. Nevertheless, more systematic experiments are required to understand the complex involvement of DNA methylation in restricting TF accessibility in ES cells and beyond.

Conclusions and Future Outlook

The genomics revolution that we have experienced in the past decade had a great influence on DNA methylation research, not only through the generation of a large panel of high-resolution resources from various cell lines and healthy as well as diseased tissues but also, to a great extent, through the novel insights that have been obtained from the genome-wide analysis of this important mark. These revealed, for example, sequences and genomic features decorated or protected from methylation, and further uncovered unexpected dynamics during cellular processes responsible for the resetting and redistribution of methylation at defined genomic sites. Furthermore, the discovery of hydroxymethylation and TET enzymes as important players in the active turnover of DNA methylation unravelled that this modification is not as inert as previously thought. It will be of great importance to understand how this cycle is regulated and its significance for gene regulation.

Genome-wide profiles further facilitated the comparison of DNA methylation patterns to other maps of chromatin modifications, transcriptional activity, and DNA sequence composition. These comparisons provided valuable information on the crosstalk to histone modifications that positively or negatively influence the deposition of DNA methylation and the dynamic interplay with TF binding at regulatory elements. While these comparative approaches were already very insightful, there is still a large amount of open questions that need to be answered

and various observations that require a more detailed exploration in order to fully understand the role of DNA methylation in genome regulation. Most notably, we still lack a comprehensive understanding on how DNA methylation regulates gene activity and how this function depends on local context. Methylation can prevent transcriptional activation through inhibiting binding of TFs to their recognition sites in regulatory sequences. This ability, however, is likely to depend on the local concentration of methylated cytosines and can be further influenced by the expression levels and DNA-binding affinities of individual or multiple TFs. The ability of methylated cytosines to inhibit TF binding could furthermore be amplified by factors that specifically recognise and bind to this modification. The key players that have been involved in the readout of methylation, the MBD proteins, still remain enigmatic in their function as translators of DNA methylation into transcriptional repressors, and more detailed *in vivo* studies are required to understand their role. The possibility that additional tissue-specific factors could exist that read DNA methylation in a sequence-specific manner and contribute to a context-dependent readout of this pervasive mark provides an attractive hypothesis that needs to be explored in more detail using genomics, biochemistry, and genetics approaches. A comprehensive understanding of the mechanisms that are influenced by DNA methylation, in combination with the detailed knowledge about the distribution and frequency of methylated cytosines along the genomes of various cell and tissue types, will allow us to generate quantitative models that could predict if the presence of DNA methylation at promoters of interest is indeed a mark of repression.

Furthermore, it remains to be fully elucidated how the deposition or removal of DNA methylation is mediated at specific genomic regions. Understanding which processes drive or block methylation to certain sites will provide valuable insight into the mechanisms that regulate the deposition of DNA methylation in a context-specific manner. Additionally, it could elucidate the order of events that distinguish if methylation acts upstream or downstream of gene activity. A careful dissection of the molecular mechanisms that regulate methylation along the genome is therefore a required step towards identifying the hierarchy of this mark in transcriptional regulation. Histone modifications such as H3K4, H3K9, or H3K36 methylation have been already implicated in repulsion or recruitment of methylation to specific genomic sites, including CGIs, repetitive elements, or transcribed gene bodies. However, there is still more to be explored, ranging from nucleosome positioning to remodelling activities of various factors, transcriptional processes, nuclear organisation, and protein–protein interactions of the writers and erasers of DNA methylation.

Acknowledgements

We apologise to all colleagues whose work we could not cite due to length limitations. We would like to thank Raffaella Santoro (DMMD, University of Zurich), Rodrigo Villasenor Molina (Baubec lab), and the anonymous reviewers for their careful reading of the manuscript and their suggestions. Research in the laboratory of T.B. is supported by the Swiss National Science Foundation (SNSF Professorship, P3_157488), the Swiss Initiative in Systems Biology (SystemsX.ch), The Novartis Research Foundation for Biomedical Research, and the University of Zurich.

Received 15 October 2016;
Received in revised form 26 January 2017;
Accepted 9 February 2017
Available online 16 February 2017

Keywords:
DNA methylation;
stem cells;
DNMT;
chromatin;
epigenetics

Abbreviations used:
mES, murine embryonic stem; DNMT, DNA methyltransferase; DNMT3A, DNA methyltransferase 3A; DNMT3B, DNA methyltransferase 3B; DNMT1, DNA methyltransferase 1; H3K36me3, histone H3 lysine-36 trimethylation; ES, embryonic stem; UHRF1, ubiquitin-like with PHD and ring finger domains 1; H3K9, histone H3 lysine 9; TET, ten-eleven translocation; 5hmC, 5-hydroxymethylcytosine; CGI, CpG island; ICM, inner cell mass; TF, transcription factor; PGC, primordial germ cell; ERV, endogenous retrovirus; MBD, methyl-CpG-binding domain; 2i, two inhibitor ES medium; H3K4me3, histone H3 lysine-4 trimethylation; TKO, Dnmt-triple-KO.

References

[1] M.M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics, Nat. Rev. Genet. 9 (2008) 465–476.

[2] A. Zemach, et al., Genome-wide evolutionary analysis of eukaryotic DNA methylation, Science 328 (2010) 916–919.

[3] R. Holliday, J.E. Pugh, DNA modification mechanisms and gene activity during development, Science 187 (1975) 226–232.

[4] A.D. Riggs, X inactivation, differentiation, and DNA methylation, Cytogenet. Cell Genet. 14 (1975) 9–25.

[5] J. Borgel, et al., Targets and dynamics of promoter DNA methylation during early mouse development, Nat. Genet. 42 (2010) 1093–1100.

[6] C. Beard, E. Li, R. Jaenisch, Loss of methylation activates Xist in somatic but not in embryonic cells, Genes Dev. 9 (1995) 2325–2334.
[7] E. Li, C. Beard, R. Jaenisch, Role for DNA methylation in genomic imprinting, Nature 366 (1993) 362–365.
[8] M. Kaneda, et al., Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting, Nature 429 (2004) 900–903.
[9] D. Jahner, et al., De novo methylation and expression of retroviral genomes during mouse embryogenesis, Nature 298 (1982) 623–628.
[10] C.P. Walsh, J.R. Chaillet, T.H. Bestor, Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, Nat. Genet. 20 (1998) 116–117.
[11] D. Bourc'his, T.H. Bestor, Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L, Nature 431 (2004) 96–99.
[12] G. Egger, et al., Epigenetics in human disease and prospects for epigenetic therapy, Nature 429 (2004) 457–463.
[13] E. Li, T.H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, Cell 69 (1992) 915–926.
[14] M. Okano, et al., DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, Cell 99 (1999) 247–257.
[15] M.G. Goll, T.H. Bestor, Eukaryotic cytosine methyltransferases, Annu. Rev. Biochem. 74 (2005) 481–514.
[16] Y. Gruenbaum, H. Cedar, A. Razin, Substrate and sequence specificity of a eukaryotic DNA methylase, Nature 295 (1982) 620–622.
[17] T.H. Bestor, V.M. Ingram, Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA, Proc. Natl. Acad. Sci. U. S. A. 80 (1983) 5559–5563.
[18] H. Leonhardt, et al., A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei, Cell 71 (1992) 865–873.
[19] G. Liang, et al., Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements, Mol. Cell. Biol. 22 (2002) 480–491.
[20] T. Chen, et al., Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b, Mol. Cell. Biol. 23 (2003) 5594–5605.
[21] M. Jackson, et al., Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells, Mol. Cell. Biol. 24 (2004) 8862–8871.
[22] J. Barau, et al., The DNA methyltransferase DNMT3C protects male germ cells from transposon activity, Science 354 (2016) 909–912.
[23] J. Du, et al., DNA methylation pathways and their crosstalk with histone methylation, Nat. Rev. Mol. Cell Biol. 16 (2015) 519–532.
[24] A. Jeltsch, R.Z. Jurkowska, Allosteric control of mammalian DNA methyltransferases—a new regulatory paradigm, Nucleic Acids Res. 44 (18) (2016) 8556–8575.
[25] I. Suetake, D. Hayata, S. Tajima, The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif, J. Biochem. 140 (2006) 763–776.
[26] J. Song, et al., Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation, Science 331 (2011) 1036–1040.
[27] S.K. Ooi, et al., DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, Nature 448 (2007) 714–717.
[28] J. Otani, et al., Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain, EMBO Rep. 10 (2009) 1235–1241.
[29] K.M. Noh, et al., Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs, Mol. Cell 59 (2015) 89–103.
[30] A. Dhayalan, et al., The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation, J. Biol. Chem. 285 (2010) 26,114–26,120.
[31] T. Baubec, et al., Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation, Nature 520 (2015) 243–247.
[32] G. Rondelet, et al., Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B, J. Struct. Biol. 194 (2016) 357–367.
[33] T. Chen, N. Tsujimoto, E. Li, The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin, Mol. Cell. Biol. 24 (2004) 9048–9058.
[34] S. Jeong, et al., Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA, Mol. Cell. Biol. 29 (2009) 5366–5376.
[35] I. Suetake, et al., Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a, Biochem. J. 437 (2011) 141–148.
[36] T. Chen, et al., A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation, J. Biol. Chem. 277 (2002) 38,746–38,754.
[37] E.S. Kang, C.W. Park, J.H. Chung, Dnmt3b, de novo DNA methyltransferase, interacts with SUMO-1 and Ubc9 through its N-terminal region and is subject to modification by SUMO-1, Biochem. Biophys. Res. Commun. 289 (2001) 862–868.
[38] B. Lee, M.T. Muller, SUMOylation enhances DNA methyltransferase 1 activity, Biochem. J. 421 (2009) 449–461.
[39] M. Bostick, et al., UHRF1 plays a role in maintaining DNA methylation in mammalian cells, Science 317 (2007) 1760–1764.
[40] K. Myant, et al., LSH and G9a/GLP complex are required for developmentally programmed DNA methylation, Genome Res. 21 (2011) 83–94.
[41] F. Chedin, M.R. Lieber, C.L. Hsieh, The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 16,916–16,921.
[42] A.A. Aravin, et al., A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice, Mol. Cell 31 (2008) 785–799.
[43] P.A. Ginno, et al., R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters, Mol. Cell 45 (2012) 814–825.

[44] A. Di Ruscio, et al., DNMT1-interacting RNAs block gene-specific DNA methylation, Nature 503 (2013) 371–376.
[45] J. Sharif, et al., The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, Nature 450 (2007) 908–912.
[46] K. Arita, et al., Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism, Nature 455 (2008) 818–821.
[47] G.V. Avvakumov, et al., Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, Nature 455 (2008) 822–825.
[48] S.B. Rothbart, et al., Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation, Nat. Struct. Mol. Biol. 19 (2012) 1155–1160.
[49] X. Liu, et al., UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9, Nat. Commun. 4 (2013) 1563.
[50] I. Suetake, et al., DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction, J. Biol. Chem. 279 (2004) 27,816–27,823.
[51] Z.X. Chen, et al., Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family, J. Cell. Biochem. 95 (2005) 902–917.
[52] C.E. Duymich, et al., DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells, Nat. Commun. 7 (2016) 11,453.
[53] D. Bourc'his, et al., Dnmt3L and the establishment of maternal genomic imprints, Science 294 (2001) 2536–2539.
[54] K. Hata, et al., Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice, Development 129 (2002) 1983–1993.
[55] N. Zamudio, et al., DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination, Genes Dev. 29 (2015) 1256–1270.
[56] M. Guenatri, et al., Plasticity in Dnmt3L-dependent and -independent modes of de novo methylation in the developing mouse embryo, Development 140 (2013) 562–572.
[57] H. Lei, et al., De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells, Development 122 (1996) 3195–3205.
[58] J. Huntriss, et al., Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells, Mol. Reprod. Dev. 67 (2004) 323–336.
[59] C.A. Gordon, S.R. Hartono, F. Chedin, Inactive DNMT3B splice variants modulate de novo DNA methylation, PLoS One 8 (2013) e69486.
[60] S.A. Smallwood, et al., Dynamic CpG island methylation landscape in oocytes and preimplantation embryos, Nat. Genet. 43 (2011) 811–814.
[61] H. Wu, Y. Zhang, Reversing DNA methylation: mechanisms, genomics, and biological functions, Cell 156 (2014) 45–68.
[62] C. Popp, et al., Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency, Nature 463 (2010) 1101–1105.
[63] P. Hajkova, et al., Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway, Science 329 (2010) 78–82.
[64] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, Science 324 (2009) 929–930.
[65] M. Tahiliani, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, Science 324 (2009) 930–935.
[66] S. Ito, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–1303.
[67] Y.F. He, et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, Science 333 (2011) 1303–1307.
[68] H. Hashimoto, et al., Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, Nucleic Acids Res. 40 (2012) 4841–4849.
[69] S. Ito, et al., Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification, Nature 466 (2010) 1129–1133.
[70] K.D. Rasmussen, K. Helin, Role of TET enzymes in DNA methylation, development, and cancer, Genes Dev. 30 (2016) 733–750.
[71] L.M. Iyer, et al., Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids, Cell Cycle 8 (2009) 1698–1710.
[72] M. Ko, et al., Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX, Nature 497 (2013) 122–126.
[73] J.H. Lee, K.S. Voo, D.G. Skalnik, Identification and characterization of the DNA binding domain of CpG-binding protein, J. Biol. Chem. 276 (2001) 44,669–44,676.
[74] H. Zhang, et al., TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine, Cell Res. 20 (2010) 1390–1393.
[75] C. Frauer, et al., Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1, PLoS One 6 (2011) e16627.
[76] Y. Xu, et al., Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development, Cell 151 (2012) 1200–1213.
[77] K. Williams, et al., TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity, Nature 473 (2011) 343–348.
[78] H. Wu, et al., Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells, Nature 473 (2011) 389–393.
[79] M.B. Stadler, et al., DNA-binding factors shape the mouse methylome at distal regulatory regions, Nature 480 (2011) 490–495.
[80] W.A. Pastor, et al., Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells, Nature 473 (2011) 394–397.
[81] H. Wu, Y. Zhang, Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation, Genes Dev. 25 (2011) 2436–2452.
[82] L. Shen, et al., Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes, Cell Stem Cell 15 (2014) 459–470.

[83] K. Moran-Crusio, et al., Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation, Cancer Cell 20 (2011) 11–24.
[84] M.M. Dawlaty, et al., Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development, Cell Stem Cell 9 (2011) 166–175.
[85] T.P. Gu, et al., The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes, Nature 477 (2011) 606–610.
[86] M.M. Dawlaty, et al., Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development, Dev. Cell 24 (2013) 310–323.
[87] M.M. Dawlaty, et al., Loss of Tet enzymes compromises proper differentiation of embryonic stem cells, Dev. Cell 29 (2014) 102–111.
[88] F. Lu, et al., Role of Tet proteins in enhancer activity and telomere elongation, Genes Dev. 28 (2014) 2103–2119.
[89] G.C. Hon, et al., 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation, Mol. Cell 56 (2014) 286–297.
[90] B.H. Ramsahoye, et al., Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 5237–5242.
[91] R. Lister, et al., Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322.
[92] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, Nat. Rev. Genet. 14 (2013) 204–220.
[93] A. Bird, et al., A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA, Cell 40 (1985) 91–99.
[94] M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes, J. Mol. Biol. 196 (1987) 261–282.
[95] M. Weber, et al., Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, Nat. Genet. 39 (2007) 457–466.
[96] H. Takeshima, et al., The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands, Genome Res. 19 (2009) 1974–1982.
[97] A.R. Krebs, et al., High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions, Elife 3 (2014) e04094.
[98] J.P. Thomson, et al., CpG islands influence chromatin structure via the CpG-binding protein Cfp1, Nature 464 (2010) 1082–1086.
[99] Y. Xu, et al., Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells, Mol. Cell 42 (2011) 451–464.
[100] J.E. Butler, J.T. Kadonaga, The RNA polymerase II core promoter: a key component in the regulation of gene expression, Genes Dev. 16 (2002) 2583–2592.
[101] A.P. Bird, DNA methylation and the frequency of CpG in animal DNA, Nucleic Acids Res. 8 (1980) 1499–1504.
[102] B.K. Duncan, J.H. Miller, Mutagenic deamination of cytosine residues in DNA, Nature 287 (1980) 560–561.
[103] J. Boyes, A. Bird, Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein, EMBO J. 11 (1992) 327–333.
[104] T. Baubec, D. Schubeler, Genomic patterns and context specific interpretation of DNA methylation, Curr. Opin. Genet. Dev. 25 (2014) 85–92.
[105] A. Molaro, et al., Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates, Cell 146 (2011) 1029–1041.
[106] E. Hodges, et al., Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment, Mol. Cell 44 (2011) 17–28.
[107] A. Meissner, et al., Genome-scale DNA methylation maps of pluripotent and differentiated cells, Nature 454 (2008) 766-U91.
[108] G.C. Hon, et al., Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues, Nat. Genet. 45 (2013) 1198–1206.
[109] P. Hajkova, et al., Epigenetic reprogramming in mouse primordial germ cells, Mech. Dev. 117 (2002) 15–23.
[110] N. Lane, et al., Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, Genesis 35 (2003) 88–93.
[111] S. Guibert, T. Forne, M. Weber, Global profiling of DNA methylation erasure in mouse primordial germ cells, Genome Res. 22 (2012) 633–641.
[112] S. Seisenberger, et al., The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells, Mol. Cell 48 (2012) 849–862.
[113] S.K. Kota, R. Feil, Epigenetic transitions in germ cell development and meiosis, Dev. Cell 19 (2010) 675–686.
[114] H. Kobayashi, et al., Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks, PLoS Genet. 8 (2012) e1002440.
[115] S.S. Hammoud, et al., Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis, Cell Stem Cell 15 (2014) 239–253.
[116] N. Kubo, et al., DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis, BMC Genomics 16 (2015) 624.
[117] H. Sasaki, Y. Matsui, Epigenetic events in mammalian germ-cell development: reprogramming and beyond, Nat. Rev. Genet. 9 (2008) 129–140.
[118] K. Shirane, et al., Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases, PLoS Genet. 9 (2013) e1003439.
[119] K.R. Stewart, et al., Dynamic changes in histone modifications precede de novo DNA methylation in oocytes, Genes Dev. 29 (2015) 2449–2462.
[120] J.A. Dahl, et al., Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition, Nature 537 (2016) 548–552.
[121] B. Zhang, et al., Allelic reprogramming of the histone modification H3K4me3 in early mammalian development, Nature 537 (2016) 553–557.
[122] Z.D. Smith, et al., A unique regulatory phase of DNA methylation in the early mammalian embryo, Nature 484 (2012) 339–344.
[123] L. Wang, et al., Programming and inheritance of parental DNA methylomes in mammals, Cell 157 (2014) 979–991.
[124] W. Mayer, et al., Demethylation of the zygotic paternal genome, Nature 403 (2000) 501–502.

[125] J. Oswald, et al., Active demethylation of the paternal genome in the mouse zygote, Curr. Biol. 10 (2000) 475–478.
[126] F. Santos, et al., Dynamic reprogramming of DNA methylation in the early mouse embryo, Dev. Biol. 241 (2002) 172–182.
[127] K. Iqbal, et al., Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3642–3647.
[128] M. Wossidlo, et al., 5-hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming, Nat. Commun. 2 (2011) 241.
[129] F. Santos, et al., Active demethylation in mouse zygotes involves cytosine deamination and base excision repair, Epigenetics Chromatin 6 (2013) 39.
[130] L.L. Carlson, A.W. Page, T.H. Bestor, Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting, Genes Dev. 6 (1992) 2536–2541.
[131] R. Hirasawa, et al., Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development, Genes Dev. 22 (2008) 1607–1616.
[132] R. Amouroux, et al., De novo DNA methylation drives 5hmC accumulation in mouse zygotes, Nat. Cell Biol. 18 (2016) 225–233.
[133] W. Reik, W. Dean, DNA methylation and mammalian epigenetics, Electrophoresis 22 (2001) 2838–2843.
[134] F. Gaudet, et al., Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing, Mol. Cell. Biol. 24 (2004) 1640–1648.
[135] H.G. Leitch, et al., Naive pluripotency is associated with global DNA hypomethylation, Nat. Struct. Mol. Biol. 20 (2013) 311–316.
[136] D. Watanabe, et al., Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis, Mech. Dev. 118 (2002) 187–190.
[137] G. Auclair, et al., Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse, Genome Biol. 15 (2014) 545.
[138] H. Wu, et al., Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes, Science 329 (2010) 444–448.
[139] G.A. Challen, et al., Dnmt3a is essential for hematopoietic stem cell differentiation, Nat. Genet. 44 (2012) 23–31.
[140] J.Y. Li, et al., Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog, Mol. Cell. Biol. 27 (2007) 8748–8759.
[141] S.D. Fouse, et al., Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation, Cell Stem Cell 2 (2008) 160–169.
[142] J.A. Hackett, et al., Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline, Development 139 (2012) 3623–3632.
[143] C.R. Farthing, et al., Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes, PLoS Genet. 4 (2008) e1000116.
[144] A.G. Smith, Embryo-derived stem cells: of mice and men, Annu. Rev. Cell Dev. Biol. 17 (2001) 435–462.
[145] Q.L. Ying, et al., The ground state of embryonic stem cell self-renewal, Nature 453 (2008) 519–523.
[146] H. Marks, et al., The transcriptional and epigenomic foundations of ground state pluripotency, Cell 149 (2012) 590–604.
[147] J. Nichols, A. Smith, Pluripotency in the embryo and in culture, Cold Spring Harb. Perspect. Biol. 4 (2012) a008128.
[148] E. Habibi, et al., Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells, Cell Stem Cell 13 (2013) 360–369.
[149] G. Ficz, et al., FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency, Cell Stem Cell 13 (2013) 351–359.
[150] M.E. Torres-Padilla, I. Chambers, Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage, Development 141 (2014) 2173–2181.
[151] A.A. Kolodziejczyk, et al., Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation, Cell Stem Cell 17 (2015) 471–485.
[152] A. Feldmann, et al., Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions, PLoS Genet. 9 (2013) e1003994.
[153] Z. Shipony, et al., Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells, Nature 513 (2014) 115–119.
[154] S.A. Smallwood, et al., Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity, Nat. Methods 11 (2014) 817–820.
[155] Z.S. Singer, et al., Dynamic heterogeneity and DNA methylation in embryonic stem cells, Mol. Cell 55 (2014) 319–331.
[156] A. Tsumura, et al., Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b, Genes Cells 11 (2006) 805–814.
[157] N. Grabole, et al., Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation, EMBO Rep. 14 (2013) 629–637.
[158] J.A. Hackett, et al., Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency, Stem Cell Rep. 1 (2013) 518–531.
[159] K. Blaschke, et al., Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells, Nature 500 (2013) 222–226.
[160] F. von Meyenn, et al., Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells, Mol. Cell 62 (2016) 983.
[161] F. Mohn, et al., Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors, Mol. Cell 30 (2008) 755–766.
[162] D.D. De Carvalho, J.S. You, P.A. Jones, DNA methylation and cellular reprogramming, Trends Cell Biol. 20 (2010) 609–617.
[163] T.S. Mikkelsen, et al., Dissecting direct reprogramming through integrative genomic analysis, Nature 454 (2008) 49–55.
[164] Y. Gao, et al., Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming, Cell Stem Cell 12 (2013) 453–469.

[165] B. Panning, R. Jaenisch, DNA hypomethylation can activate Xist expression and silence X-linked genes, Genes Dev. 10 (1996) 1991–2002.
[166] M.M. Karimi, et al., DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs, Cell Stem Cell 8 (2011) 676–687.
[167] J.E. Dodge, et al., Histone H3-K9 methyltransferase ESET is essential for early development, Mol. Cell. Biol. 24 (2004) 2478–2486.
[168] H.M. Rowe, et al., KAP1 controls endogenous retroviruses in embryonic stem cells, Nature 463 (2010) 237–240.
[169] T. Matsui, et al., Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET, Nature 464 (2010) 927–931.
[170] M. Walter, et al., An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells, Elife 5 (2016) http://dx.doi.org/10.7554/eLife.11418.
[171] E.H. Davidson, Emerging properties of animal gene regulatory networks, Nature 468 (2010) 911–920.
[172] C.D. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control, Nat. Rev. Genet. 17 (2016) 487–500.
[173] M. Sakaue, et al., DNA methylation is dispensable for the growth and survival of the extraembryonic lineages, Curr. Biol. 20 (2010) 1452–1457.
[174] P.H. Tate, A.P. Bird, Effects of DNA methylation on DNA-binding proteins and gene expression, Curr. Opin. Genet. Dev. 3 (1993) 226–231.
[175] A.C. Bell, G. Felsenfeld, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene, Nature 405 (2000) 482–485.
[176] A.T. Hark, et al., CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus, Nature 405 (2000) 486–489.
[177] S. Domcke, et al., Competition between DNA methylation and transcription factors determines binding of NRF1, Nature 528 (2015) 575–579.
[178] C.G. Spruijt, et al., Dynamic readers for 5-(hydroxy)-methylcytosine and its oxidized derivatives, Cell 152 (2013) 1146–1159.
[179] S. Hu, et al., DNA methylation presents distinct binding sites for human transcription factors, Elife 2 (2013) e00726.
[180] H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation, Nat. Rev. Genet. 17 (2016) 551–565.
[181] R.R. Meehan, et al., Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs, Cell 58 (1989) 499–507.
[182] B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins, Mol. Cell. Biol. 18 (1998) 6538–6547.
[183] X. Nan, et al., Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex, Nature 393 (1998) 386–389.
[184] T. Baubec, et al., Methylation-dependent and -independent genomic targeting principles of the MBD protein family, Cell 153 (2013) 480–492.
[185] M.J. Lyst, et al., Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor, Nat. Neurosci. 16 (2013) 898–902.
[186] X. Nan, F.J. Campoy, A. Bird, MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin, Cell 88 (1997) 471–481.
[187] B. Hendrich, et al., Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development, Genes Dev. 15 (2001) 710–723.
[188] I. Martin Caballero, et al., The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation, PLoS One 4 (2009) e4315.
[189] G.L. Xu, T.H. Bestor, Cytosine methylation targetted to pre-determined sequences, Nat. Genet. 17 (1997) 376–378.
[190] L.A. Gilbert, et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, Cell 154 (2013) 442–451.
[191] B. Chen, et al., Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system, Cell 155 (2013) 1479–1491.
[192] X.S. Liu, et al., Editing DNA methylation in the mammalian genome, Cell 167 (2016) 233–247 e17.
[193] A. Amabile, et al., Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing, Cell 167 (2016) 219–232 e14.
[194] F. Lienert, et al., Identification of genetic elements that autonomously determine DNA methylation states, Nat. Genet. 43 (2011) 1091–1097.
[195] R.E. Thurman, et al., The accessible chromatin landscape of the human genome, Nature 489 (2012) 75–82.
[196] H. Stroud, et al., 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells, Genome Biol. 12 (2011) R54.
[197] M.T. Maurano, et al., Large-scale identification of sequence variants influencing human transcription factor occupancy *in vivo*, Nat. Genet. 47 (2015) 1393–1401.

7.2. Published Chapter on Genome-Wide Profiling of DNA Methyltransferases in Mammalian Cells

Authors: Massimiliano Manzo, **Christina Ambrosi**, and Tuncay Baubec.

Published: CpG Islands, Methods in Molecular Biology, 2018; doi: 10.1016/j.jmb.2017.02.008
https://link.springer.com/protocol/10.1007%2F978-1-4939-7768-0_9

Summary: Here, we describe a method to investigate the binding preferences of mammalian DNA methyltransferases (DNMT). Using prior biotin-tagging of DNMTs in engineered cells, stringent ChIP-seq based on biotin-avidin interactions circumvents limitations arising from the lack of high specificity of commercially available antibodies. This ensures reproducible results in a high-throughput manner. Follow-up bioinformatic analyses allow specific and reliable determination of DNMT genome-wide binding.

Author contribution: I participated in the writing and figure preparation. I contributed to the editing and revision of this book chapter.

Chapter 9

Genome-Wide Profiling of DNA Methyltransferases in Mammalian Cells

Massimiliano Manzo, Christina Ambrosi, and Tuncay Baubec

Abstract

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is currently the method of choice to determine binding sites of chromatin-associated factors in a genome-wide manner. Here, we describe a method to investigate the binding preferences of mammalian DNA methyltransferases (DNMT) based on ChIP-seq using biotin-tagging. Stringent ChIP of DNMT proteins based on the strong interaction between biotin and avidin circumvents limitations arising from low antibody specificity and ensures reproducible enrichment. DNMT-bound DNA fragments are ligated to sequencing adaptors, amplified and sequenced on a high-throughput sequencing instrument. Bioinformatic analysis gives valuable information about the binding preferences of DNMTs genome-wide and around promoter regions. This method is unconventional due to the use of genetically engineered cells; however, it allows specific and reliable determination of DNMT binding.

Key words ChIP-seq, Immunoprecipitation, In vivo biotinylation, Next-generation sequencing, DNA methyltransferases, CpG islands

1 Introduction

Methylation of cytosine bases is one of the best mechanistically understood epigenetic modifications and plays various roles in genome regulation. Three conserved enzymes are responsible for the deposition of methyl groups to cytosine bases in mammals—the de novo DNA methyltransferases 3A (DNMT3A) and DNMT3B, as well as the maintenance DNA methyltransferase 1 (DNMT1) [1, 2]. In mammals, DNA methylation occurs at the majority of CpG dinucleotides throughout the entire genome, only CpG islands remain largely protected from methylation [3]. Based on biochemical studies, the mechanism of DNA methylation has been largely elucidated, but how DNA methylation patterns are precisely set along the genome and to what extent these cause further regulation remains to be understood in full detail.

Tanya Vavouri and Miguel A. Peinado (eds.), *CpG Islands: Methods and Protocols*, Methods in Molecular Biology, vol. 1766, https://doi.org/10.1007/978-1-4939-7768-0_9,

8. Contributions

8.1. ChromID identifies the protein interactome at chromatin marks

Authors: Rodrigo Villaseñor, Ramon Pfaendler, **Christina Ambrosi**, Stefan Butz, Sara Giuliani, Elana Bryan, Thomas W. Sheahan, Annika L. Gable, Nina Schmolka, Massimiliano Manzo, Joël Wirz, Christian Feller, Christian von Mering, Ruedi Aebersold, Philipp Voigt and Tuncay Baubec.

Published: Nature Biotechnology, 2020; doi: 10.1038/s41587-020-0434-2
https://www.nature.com/articles/s41587-020-0434-2

Summary: This part contains the published article titled *ChromID identifies the protein interactome at chromatin marks*. Here, we describe ChromID, a new tool to attain a detailed overview of chromatin-dependent protein interaction networks using engineered chromatin readers (eCRs) in combination with a biotin ligase BASU for proximity biotinylation. In particular, we uncovered the interactome of bivalent promoters carrying the histone modifications H3K4me3 and H3K27me3.

Author contribution: I performed ChIP-seq experiments for trimethylation of lysine 4, 27, and 9 on histone 3 in murine embryonic stem (mES) cells. Obtained data was used in Figure 1 D and G, 2 A-C, 4 A-C, 5 B and Supplementary Figure S4, S5, S7 D-E, S9 C-D, S11 D-F, S12, and S14 B. I further generated an *Eed-/-* mES cell line via CRISPR-Cas9 and conducted immunoblotting for total histone 1, histone 3, and trimethylation of lysine 27 on histone 3 to control for a depletion of the named chromatin mark in these knock-out cells, seen in Supplementary Figure 6E. Additionally, I performed PolyA-RNA-seq experiments in various mES cell lines containing engineered chromatin readers (not included in the final manuscript). I contributed to the writing of the methods part and to the revisions.

Number of Citations: 1

nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-020-0434-2

ChromID identifies the protein interactome at chromatin marks

Rodrigo Villaseñor[1], Ramon Pfaendler[1], Christina Ambrosi[1,2], Stefan Butz[1,2], Sara Giuliani[1], Elana Bryan[3], Thomas W. Sheahan[3], Annika L. Gable[2,4], Nina Schmolka[1], Massimiliano Manzo[1,2], Joël Wirz[1], Christian Feller[5], Christian von Mering[4], Ruedi Aebersold[5,6], Philipp Voigt[3] and Tuncay Baubec[1]✉

Chromatin modifications regulate genome function by recruiting proteins to the genome. However, the protein composition at distinct chromatin modifications has yet to be fully characterized. In this study, we used natural protein domains as modular building blocks to develop engineered chromatin readers (eCRs) selective for DNA methylation and histone tri-methylation at H3K4, H3K9 and H3K27 residues. We first demonstrated their utility as selective chromatin binders in living cells by stably expressing eCRs in mouse embryonic stem cells and measuring their subnuclear localization, genomic distribution and histone-modification-binding preference. By fusing eCRs to the biotin ligase BASU, we established ChromID, a method for identifying the chromatin-dependent protein interactome on the basis of proximity biotinylation, and applied it to distinct chromatin modifications in mouse stem cells. Using a synthetic dual-modification reader, we also uncovered the protein composition at bivalently modified promoters marked by H3K4me3 and H3K27me3. These results highlight the ability of ChromID to obtain a detailed view of protein interaction networks on chromatin.

Chromatin and many chemical modifications on histones and DNA play critical roles in organismal development and human health[1]. These modifications are recognized by specialized reader domains in regulatory proteins and multiprotein complexes[2,3]. Depending on the presence and composition of modifications at genomic sites, regulatory factors can associate with chromatin in a spatiotemporal manner[4]. However, a major challenge in the field remains understanding how this chemical language on chromatin defines the protein interactome of the genome.

In recent years, proteomics-based assays have helped measure the affinity of proteins for chromatin marks. Current methods probe the cellular proteome using synthetic histone peptides, methylated DNA probes or in vitro-reconstituted nucleosomes[5–9]. In addition, proteins bound to specific genomic segments can be identified using enrichment via antibodies, DNA-sequence-specific probes or, more recently, engineered dCas9 fusion proteins[10–14]. Although these studies have greatly enhanced current knowledge about interactions between proteins and chromatin marks, the available methods rely on artificial chromatin, protein–protein cross-linking or methods that require access to the underlying DNA, leading to chromatin disruption. Therefore, new approaches are required that enable detection of dynamic interactions between proteins and physiological chromatin in living cells.

In this study, we developed ChromID to identify the local protein composition at individual and combinatorial chromatin marks. To this end, we used the reader domains of well-established chromatin regulators as modules to build eCRs. We first quantified and functionally validated the genome-wide binding and/or histone post-translational modification (PTM) interaction preferences of individual eCRs toward DNA methylation, H3K9me3, H3K4me3 and H3K27me3, demonstrating their applicability as selective binders in mouse stem cells. Finally, we used the specificity of eCRs to recruit promiscuous biotin ligases to detect proteins associated with these individual chromatin modifications in mouse embryonic stem cells (mESCs), revealing similarities and differences in the protein composition between these marks. By coupling ChromID to a synthetic dual-modification reader, we also detected proteins associated with genomic regions marked bivalently with H3K4me3 and H3K27me3 modifications.

Results

Generation and characterization of eCRs in mESCs. We first assembled well-characterized chromatin reader domains into synthetic reporter proteins to test their affinity and specificity for individual chromatin modifications in living cells. We used the chromodomains specific for H3K27me3 from CBX7 and *Drosophila* Polycomb (dPC)[15,16], the H3K9me3-specific chromodomain from CBX1 (refs. [17,18]), the Phd domain specific for H3K4me3 from TAF3 (ref. [19]) and the MBD domains from the DNA methylation readers MBD1 and MeCP2 (refs. [20,21]) (Fig. 1a). cDNA sequences were assembled as either single- or dual-domain constructs in a protein expression cassette containing a biotin acceptor site for biochemical purification, a nuclear localization signal (NLS) and enhanced green fluorescent protein (eGFP) for live-cell imaging and detection (Fig. 1a and Supplementary Fig. 1a). All constructs were integrated into a defined site in the mouse genome via recombinase-mediated cassette exchange (RMCE)[22], enabling fast generation of mESC lines stably expressing the proteins from the same genomic location and under the control of the same promoter (Fig. 1b). Measurements of eGFP fluorescence and protein levels indicated that all generated cell

[1]Department of Molecular Mechanism of Disease, University of Zurich, Zurich, Switzerland. [2]Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland. [3]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. [4]Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland. [5]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. [6]Faculty of Science, University of Zurich, Zurich, Switzerland. ✉e-mail: tuncay.baubec@uzh.ch

8.2. Efficient Pre-mRNA Cleavage Prevents Replication Stress-Associated Genome Instability

Authors: Federico Teloni, Jone Michelena, Aleksandra Lezaja, Sinan Kilic, **Christina Ambrosi**, Shruti Menon, Jana Dobrovolna, Ralph Imhof, Pavel Janscak, Tuncay Baubec, and Matthias Altmeyer.

Published: Molecular Cell, 2019; doi: 10.1016/j.molcel.2018.11.036
https://www.sciencedirect.com/science/article/pii/S1097276518310062?via%3Dihub

Summary: Here, my contribution to published article titled *Efficient Pre-mRNA Cleavage Prevents Replication Stress-Associated Genome Instability* is described. Through a multi-screening strategy we identified the polyadenylation-complex member WDR33 to allow replication stress resilience. We described that impaired pre-mRNA cleavage leads to replication fork stalling, elevated origin firing and subsequent replication stress-associated genome instability.

Author contribution: I performed and analysed ChIP-qPCR and -seq experiments for RNA-Polymerase II phosphorylated on serine 2 in siControl and siWDR33 knock-down U2OS cell lines. ChIP-seq data analysis is shown in Figure 5 A-C. ChIP-qPCR experiments were not included in the final manuscript. I contributed to the completion of methods part and to the revisions.

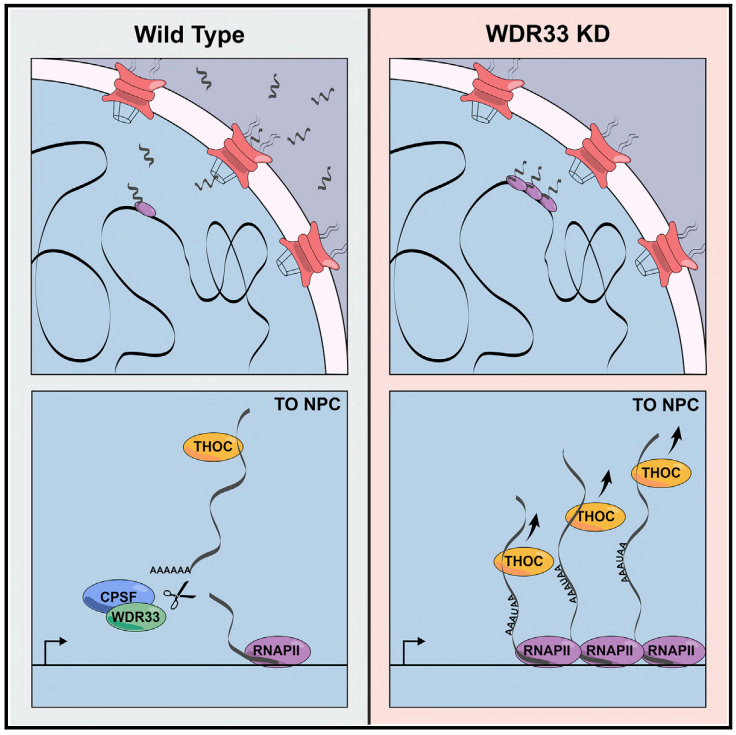
Number of Citations: 10

Article

Molecular Cell

Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability

Graphical Abstract



Highlights

- A convergent multi-screening approach reveals modulators of replication stress (RS)
- Efficient WDR33-mediated pre-mRNA cleavage confers RS resilience
- RNA:DNA-hybrid formation occurs upon RS-induced DNA breakage
- THO nuclear export complex drives RS sensitivity of WDR33-depleted cells

Authors

Federico Teloni, Jone Michelena, Aleksandra Lezaja, ..., Pavel Janscak, Tuncay Baubec, Matthias Altmeyer

Correspondence

matthias.altmeyer@uzh.ch

In Brief

Replication stress is a hallmark of many cancers. Teloni et al. identify the pre-mRNA cleavage factor WDR33 as regulator of replication stress resilience and demonstrate that, when WDR33 function is impaired, unreleased nascent transcripts and genomic loci re-localize toward the nuclear periphery, where they cause replication stress and DNA damage.

Teloni et al., 2019, Molecular Cell *73*, 670–683
February 21, 2019
https://doi.org/10.1016/j.molcel.2018.11.036

CellPress

8.3. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands

Authors: Massimiliano Manzo, Joël Wirz, Christina Ambrosi, Rodrigo Villaseñor, Bernd Roschitzki and Tuncay Baubec.

Published: EMBO, 2017; doi: 10.15252/embj.201797038
https://www.embopress.org/doi/full/10.15252/embj.201797038

Summary: This part comprises my contribution to the published article entitled *Isoform-specific localisation of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands*. Here, we studied the individual genomic binding of DNMT3A isoforms in mES cells as well as NPCs and identified that the longer isoform DNMT3A1 localises to and *de novo* - methylates CpG island shores of Polycomb-regulated, bivalent promoters of developmental genes in order to regulate the turnover of 5mC at these sites.

Author contribution: I performed RT-qPCR experiments for gene expression analysis of *Dnmt3* genes and *Hprt* housekeeping control in mES cells, seen in Appendix Figure S1 B, 10 C. I helped to extract genomic DNA from various mES cell lines for subsequent DNA methylation level measurements by mass spectrometry (Supplementary Figure S1 C). I also contributed to the writing of methods part and the revision of the manuscript.

Number of Citations: 37

Published online: October 26, 2017

Article

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands

Massimiliano Manzo[1,2], Joël Wirz[1], Christina Ambrosi[1,2], Rodrigo Villaseñor[1], Bernd Roschitzki[3] & Tuncay Baubec[1,*]

Abstract

DNA methylation is a prevalent epigenetic modification involved in transcriptional regulation and essential for mammalian development. While the genome-wide distribution of this mark has been studied to great detail, the mechanisms responsible for its correct deposition, as well as the cause for its aberrant localization in cancers, have not been fully elucidated. Here, we have compared the activity of individual DNMT3A isoforms in mouse embryonic stem and neuronal progenitor cells and report that these isoforms differ in their genomic binding and DNA methylation activity at regulatory sites. We identify that the longer isoform DNMT3A1 preferentially localizes to the methylated shores of bivalent CpG island promoters in a tissue-specific manner. The isoform-specific targeting of DNMT3A1 coincides with elevated hydroxymethylcytosine (5-hmC) deposition, suggesting an involvement of this isoform in mediating turnover of DNA methylation at these sites. Through genetic deletion and rescue experiments, we demonstrate that this isoform-specific recruitment plays a role in *de novo* DNA methylation at CpG island shores, with potential implications on H3K27me3-mediated regulation of developmental genes.

Keywords CpG islands; DNA methylation; DNMT3A; H3K27me3; Polycomb
Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Transcription
DOI 10.15252/embj.201797038 | Received 29 March 2017 | Revised 1 October 2017 | Accepted 3 October 2017

Introduction

DNA methylation is a well-established epigenetic mark involved in gene regulation and genome stability. The importance of DNA methylation for mammalian genome function is apparent by the lethal phenotypes observed upon individual and combined knockout of DNA methyltransferases, whereas DNMT1 and DNMT3B knock-outs are embryonic lethal and DNMT3A knock-outs are lethal 4 weeks after birth (Lei *et al*, 1996; Okano *et al*, 1999; Chen *et al*, 2003). Furthermore, aberrant deposition of DNA methylation is frequently observed in cancers (Baylin, 2005; Yan *et al*, 2011). Recent genome-wide initiatives explored the distribution of methylated CpGs in various cell types and tissues at single-base pair resolution (Lister *et al*, 2009; Stadler *et al*, 2011; Hon *et al*, 2013). These datasets were crucial in identifying the frequency and localization of methylated cytosines, and also to monitor changes in methylation during cellular transitions, including differentiation in healthy individuals (Schultz *et al*, 2015) or cellular transformation (Akalin *et al*, 2012; Hovestadt *et al*, 2014). Furthermore, the discovery of 5-hydroxymethylcytosine (5-hmC) as an additional modification of mammalian genomes and the characterization of TET-mediated DNA methylation removal (Kriaucionis & Heintz, 2009; Tahiliani *et al*, 2009; Kohli & Zhang, 2013) provide compelling evidence that DNA methylation is highly dynamic and undergoes constant turnover at regulatory sites (Stroud *et al*, 2011; Feldmann *et al*, 2013; Kohli & Zhang, 2013).

The *de novo* enzymes DNMT3A and DNMT3B are responsible for establishing DNA methylation, while DNMT1 is the maintenance methyltransferase responsible for propagation of this mark at CpG dinucleotides after DNA replication. In addition, the *de novo* DNMTs are responsible for the frequently occurring non-CpG methylation in mammalian genomes and contribute to maintenance of CpG methylation through filling up gaps after DNMT1 or counteracting active demethylation (Ramsahoye *et al*, 2000; Liang *et al*, 2002; Jackson *et al*, 2004; Arand *et al*, 2012). Recently, DNMT3C, a novel rodent-specific member of the *de novo* DNMT family, has been identified to regulate DNA methylation in the male germline (Barau *et al*, 2016). How DNMTs are correctly recruited to the genome in order to establish and maintain DNA methylation is not completely understood. In addition, splicing and alternative promoter usage gives rise to various catalytically active and inactive DNMT isoforms with tissue- and cancer-specific expression preferences (Chen *et al*, 2002; La Salle & Trasler, 2006; Gopalakrishnan *et al*, 2009; Duymich *et al*, 2016), revealing a complex regulation of DNA methylation through isoform variation. Previous studies have measured subcellular localization, catalytic activity, and targeting specificities of these isoforms in various cell types and *in vitro* (Chen *et al*, 2002, 2003;

1 Department of Molecular Mechanisms of Disease, University of Zurich, Zurich, Switzerland
2 Molecular Life Sciences PhD Program of the Life Sciences Zurich Graduate School, University of Zurich, Zurich, Switzerland
3 Functional Genomics Center Zurich, ETH and University of Zurich, Zurich, Switzerland
*Corresponding author. Tel: +41 44 635 5438; E-mail: tuncay.baubec@uzh.ch

9. References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell *13*, 67-76.

Alberts, B.J., A.; Lewis, A.; Raff, M.; Roberts, K.; Walter, P. (2002). Molecular Biology of the Cell, 4th edition (New York: Garland Science).

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.

Altmeyer, M., Toledo, L., Gudjonsson, T., Grofte, M., Rask, M.B., Lukas, C., Akimov, V., Blagoev, B., Bartek, J., and Lukas, J. (2013). The chromatin scaffold protein SAFB1 renders chromatin permissive for DNA damage signaling. Mol Cell *52*, 206-220.

Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. J Mol Biol *429*, 1459-1475.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P.*, et al.* (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat Struct Mol Biol *21*, 366-374.

Balazsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910-925.

Balbo, P.B., and Bohm, A. (2007). Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. Structure *15*, 1117-1131.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem *280*, 17732-17736.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183-193.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartkowiak, B., and Greenleaf, A.L. (2011). Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? Transcription *2*, 115-119.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev *24*, 2303-2316.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K.*, et al.* (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707-719.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-247.

Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480-492.

Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H.*, et al.* (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat Genet *45*, 1479-1482.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321-1325.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet *15*, 163-175.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bertero, A., Brown, S., Madrigal, P., Osnato, A., Ortmann, D., Yiangou, L., Kadiwala, J., Hubner, N.C., de Los Mozos, I.R., Sadee, C.*, et al.* (2018). The SMAD2/3 interactome reveals that TGFbeta controls m(6)A mRNA methylation in pluripotency. Nature *555*, 256-259.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nat Protoc *2*, 1034-1043.

Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. EMBO J *12*, 585-594.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res *8*, 1499-1504.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell *25*, 309-322.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell *38*, 179-190.

Blank, A., Gallant, J.A., Burgess, R.R., and Loeb, L.A. (1986). An RNA polymerase mutant with reduced accuracy of chain elongation. Biochemistry *25*, 5920-5928.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev *20*, 1123-1136.

Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., Zhou, S., King, D., Shen, P.S., Weibezahn, J.*, et al.* (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell *151*, 1042-1054.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L.*, et al.* (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol *19*, 1273-1281.

Brueckner, F., and Cramer, P. (2007). DNA photodamage recognition by RNA polymerase II. FEBS Lett *581*, 2757-2760.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol *18*, 299-314.

Busskamp, V., Lewis, N.E., Guye, P., Ng, A.H., Shipman, S.L., Byrne, S.M., Sanjana, N.E., Murn, J., Li, Y., Li, S.*, et al.* (2014). Rapid neurogenesis through transcriptional activation in human stem cells. Mol Syst Biol *10*, 760.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D.*, et al.* (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Mol Cell *49*, 571-582.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193-198.

Cao, Z., and Grima, R. (2020). Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proc Natl Acad Sci U S A *117*, 4682-4692.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P*., et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Carvalho, S., Raposo, A.C., Martins, F.B., Grosso, A.R., Sridhara, S.C., Rino, J., Carmo-Fonseca, M., and de Almeida, S.F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res *41*, 2881-2893.

Carvalho, S., Vitor, A.C., Sridhara, S.C., Martins, F.B., Raposo, A.C., Desterro, J.M., Ferreira, J., and de Almeida, S.F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Elife *3*, e02482.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells *26*, 1496-1505.

Chan, R.T., Peters, J.K., Robart, A.R., Wiryaman, T., Rajashankar, K.R., and Toor, N. (2018). Structural basis for the second step of group II intron splicing. Nat Commun *9*, 4676.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol *18*, 495-506.

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., and Cao, X. (2017). Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. Cell *170*, 492-506 e414.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Chen, R., Zhao, W.Q., Fang, C., Yang, X., and Ji, M. (2020). Histone methyltransferase SETD2: a potential tumor suppressor in solid cancers. J Cancer *11*, 3349-3356.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem *282*, 21901-21912.

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO Rep *6*, 354-360.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol *18*, 407-422.

Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., and Jinek, M. (2018). Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat Struct Mol Biol *25*, 135-138.

Cortez, D., Wang, Y., Qin, J., and Elledge, S.J. (1999). Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. Science *286*, 1162-1166.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cramer, P. (2004). Structure and function of RNA polymerase II. Adv Protein Chem *67*, 1-42.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Czudnochowski, N., Bosken, C.A., and Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat Commun *3*, 842.

D'Alessandro, G., and d'Adda di Fagagna, F. (2017). Transcription and DNA Damage: Holding Hands or Crossing Swords? J Mol Biol *429*, 3215-3229.

Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C.*, et al.* (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature *463*, 360-363.

Dasgupta, M., Dermawan, J.K., Willard, B., and Stark, G.R. (2015). STAT3-driven transcription depends upon the dimethylation of K49 by EZH2. Proc Natl Acad Sci U S A *112*, 3985-3990.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sorensen, C.S., Petersen, N.H., Sorensen, P.H., Lukas, C., Bartek, J.*, et al.* (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat Struct Mol Biol *19*, 803-810.

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., and Muegge, K. (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol *8*, 1448-1454.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and consequences of alternative polyadenylation. Mol Cell *43*, 853-866.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol *6*, 227.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol Cell *62*, 668-680.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561-564.

Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schubeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575-579.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Dreyfus, M., and Regnier, P. (2002). The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. Cell *111*, 611-613.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat Commun *7*, 12626.

Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res *70*, 4287-4291.

Eberharter, A., and Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Rep *3*, 224-229.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719-2729.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J *27*, 406-420.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457-463.

Ellis, R.J. (2006). Molecular chaperones: assisting assembly in addition to folding. Trends Biochem Sci *31*, 395-401.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T., Tkac, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D.*, et al.* (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Mol Cell *49*, 872-883.

Fahey, C.C., and Davis, I.J. (2017). SETting the Stage for Cancer Development: SETD2 and the Consequences of Lost Methylation. Cold Spring Harb Perspect Med *7*.

Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature *434*, 605-611.

Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., Kaper, F., Fan, J.B., Zhang, K., Chun, J.*, et al.* (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods *13*, 241-244.

Fang, D., Gan, H., Lee, J.H., Han, J., Wang, Z., Riester, S.M., Jin, L., Chen, J., Zhou, H., Wang, J.*, et al.* (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science *352*, 1344-1348.

Fant, C.B., Levandowski, C.B., Gupta, K., Maas, Z.L., Moir, J., Rubin, J.D., Sawyer, A., Esbin, M.N., Rimel, J.K., Luyties, O.*, et al.* (2020). TFIID Enables RNA Polymerase II Promoter-Proximal Pausing. Mol Cell.

Faure, A.J., Schmiedel, J.M., and Lehner, B. (2017). Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. Cell Syst *5*, 471-484 e474.

Feng, Y.Q., Seibler, J., Alami, R., Eisen, A., Westerman, K.A., Leboulch, P., Fiering, S., and Bouhassira, E.E. (1999). Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J Mol Biol *292*, 779-785.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stutzer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell *53*, 49-62.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J.*, et al.* (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351-359.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116-1122.

Flemr, M., and Buhler, M. (2015). Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. Cell Rep *12*, 709-716.

Fousteri, M., and Mullenders, L.H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res *18*, 73-84.

Fradet-Turcotte, A., Canny, M.D., Escribano-Diaz, C., Orthwein, A., Leung, C.C., Huang, H., Landry, M.C., Kitevski-LeBlanc, J., Noordermeer, S.M., Sicheri, F.*, et al.* (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. Nature *499*, 50-54.

Frietze, S., and Farnham, P.J. (2011). Transcription factor effector domains. Subcell Biochem *52*, 261-277.

Fujisawa, T., and Filippakopoulos, P. (2017). Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. Nat Rev Mol Cell Biol *18*, 246-262.

Fujita, K., Iwaki, M., and Yanagida, T. (2016). Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nat Commun *7*, 13788.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res *31*, 2305-2312.

Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. Genome Biol *10*, R29.

Gansen, A., Toth, K., Schwarz, N., and Langowski, J. (2015). Opposing roles of H3- and H4-acetylation in the regulation of nucleosome structure--a FRET study. Nucleic Acids Res *43*, 1433-1443.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell *45*, 344-356.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol *8*, 113-126.

Gehre, M., Bunina, D., Sidoli, S., Lubke, M.J., Diaz, N., Trovato, M., Garcia, B.A., Zaugg, J.B., and Noh, K.M. (2020). Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. Nat Genet *52*, 273-282.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Gong, F., Chiu, L.Y., Cox, B., Aymard, F., Clouaire, T., Leung, J.W., Cammarata, M., Perez, M., Agarwal, P., Brodbelt, J.S.*, et al.* (2015). Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. Genes Dev *29*, 197-211.

Graham, T.G., Walter, J.C., and Loparo, J.J. (2016). Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. Mol Cell *61*, 850-858.

Grun, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat Methods *11*, 637-640.

Hacker, K.E., Fahey, C.C., Shinsky, S.A., Chiang, Y.J., DiFiore, J.V., Jha, D.K., Vo, A.H., Shavit, J.A., Davis, I.J., Strahl, B.D.*, et al.* (2016). Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation. J Biol Chem *291*, 21283-21295.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hamperl, S., and Cimprich, K.A. (2016). Conflict Resolution in the Genome: How Transcription and Replication Make It Work. Cell *167*, 1455-1467.

Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol *9*, 958-970.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291-1300.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res *40*, 4841-4849.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O.*, et al.* (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol *17*, 77.

Hebenstreit, D. (2013). Are gene loops the cause of transcriptional noise? Trends Genet *29*, 333-338.

Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta *1829*, 55-62.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev *15*, 1593-1612.

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nunez, J., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*, 649-662.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Ho, T.H., Park, I.Y., Zhao, H., Tong, P., Champion, M.D., Yan, H., Monzon, F.A., Hoang, A., Tamboli, P., Parker, A.S.*, et al.* (2016). High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. Oncogene *35*, 1565-1574.

Hoffman, E.A., Frey, B.L., Smith, L.M., and Auble, D.T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. J Biol Chem *290*, 26404-26411.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

Hu, M., Sun, X.J., Zhang, Y.L., Kuang, Y., Hu, C.Q., Wu, W.L., Shen, S.H., Du, T.T., Li, H., He, F.*, et al.* (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. Proc Natl Acad Sci U S A *107*, 2956-2961.

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B.S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F.*, et al.* (2019). Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. Nature *567*, 414-419.

Huang, Y., Gu, L., and Li, G.M. (2018). H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. J Biol Chem *293*, 7811-7823.

Huh, D., and Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. Nat Genet *43*, 95-100.

Huh, I., Zeng, J., Park, T., and Yi, S.V. (2013). DNA methylation and transcriptional noise. Epigenetics Chromatin *6*, 9.

Husmann, D., and Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. Nat Struct Mol Biol *26*, 880-889.

Hustedt, N., and Durocher, D. (2016). The control of DNA repair by the cell cycle. Nat Cell Biol *19*, 1-9.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet *6*, e1001134.

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017). Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature *547*, 419-424.

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol *24*, 8862-8871.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet *10*, 833-844.

Jambhekar, A., Dhall, A., and Shi, Y. (2019). Roles and regulation of histone methylation in animal development. Nat Rev Mol Cell Biol *20*, 625-641.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol Cell *52*, 473-484.

Joazeiro, C.A.P. (2019). Mechanisms and functions of ribosome-associated protein quality control. Nat Rev Mol Cell Biol *20*, 368-383.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384-388.

Jones, M.J., Goodman, S.J., and Kobor, M.S. (2015). DNA methylation and healthy human aging. Aging Cell *14*, 924-932.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167-177.

Kanu, N., Gronroos, E., Martinez, P., Burrell, R.A., Yi Goh, X., Bartkova, J., Maya-Mendoza, A., Mistrik, M., Rowan, A.J., Patel, H.*, et al.* (2015). SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. Oncogene *34*, 5699-5708.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kar, G., Kim, J.K., Kolodziejczyk, A.A., Natarajan, K.N., Torlai Triglia, E., Mifsud, B., Elderkin, S., Marioni, J.C., Pombo, A., and Teichmann, S.A. (2017). Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat Commun *8*, 36.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M.*, et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Kim, T.W., Kang, B.H., Jang, H., Kwak, S., Shin, J., Kim, H., Lee, S.E., Lee, S.M., Lee, J.H., Kim, J.H.*, et al.* (2015). Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates

PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes During Exit from Pluripotency. Stem Cells *33*, 2442-2455.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol *25*, 3305-3316.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet *41*, 376-381.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865-868.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol *11*, 363-371.

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012). Homologous recombination and its regulation. Nucleic Acids Res *40*, 5795-5818.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol *23*, 4207-4218.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A.*, et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97.

Kumar, N., Singh, A., and Kulkarni, R.V. (2015). Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. PLoS Comput Biol *11*, e1004292.

Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A.*, et al.* (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol Cell *44*, 609-620.

Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol *20*, 406-420.

Kustatscher, G., Wills, K.L., Furlan, C., and Rappsilber, J. (2014). Chromatin enrichment for proteomics. Nat Protoc *9*, 2090-2099.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950-953.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A.*, et al.* (2018). RNA velocity of single cells. Nature *560*, 494-498.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet *32*, 42-56.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495-501.

Lee, C.H., Wu, J., and Li, B. (2013). Chromatin remodelers fine-tune H3K36me-directed deacetylation of neighbor nucleosomes by Rpd3S. Mol Cell *52*, 255-263.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *278*, 8897-8903.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet *3*, 662-673.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell *153*, 590-600.

Li, G.M. (2008). Mechanisms and functions of DNA mismatch repair. Cell Res *18*, 85-98.

Liu, H., Jin, T., Guan, J., and Zhou, S. (2014). Histone modifications involved in cassette exon inclusions: a quantitative and interpretable analysis. BMC Genomics *15*, 1148.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun *4*, 1563.

Ljungman, M., and Lane, D.P. (2004). Transcription - guarding the genome by sensing DNA damage. Nat Rev Cancer *4*, 727-737.

Lord, C.J., and Ashworth, A. (2012). The DNA damage response and cancer therapy. Nature *481*, 287-294.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luco, R.F., and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr Opin Genet Dev *21*, 366-372.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996-1000.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luijsterburg, M.S., Acs, K., Ackermann, L., Wiegant, W.W., Bekker-Jensen, S., Larsen, D.H., Khanna, K.K., van Attikum, H., Mailand, N., and Dantuma, N.P. (2012). A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure. EMBO J *31*, 2511-2527.

Lun, A.T., and Smyth, G.K. (2014). De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. Nucleic Acids Res *42*, e95.

Luo, W., and Bentley, D. (2004). A ribonucleolytic rat torpedoes RNA polymerase II. Cell *119*, 911-914.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev *20*, 954-965.

Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol Cell Biol *32*, 2608-2617.

Lutz, C.S. (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. ACS Chem Biol *3*, 609-617.

Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L.*, et al.* (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J *31*, 317-329.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887-900.

Malys, N., and McCarthy, J.E. (2011). Translation initiation: variations in the mechanism can be anticipated. Cell Mol Life Sci *68*, 991-1003.

Manzo, M., Ambrosi, C., and Baubec, T. (2018). Genome-Wide Profiling of DNA Methyltransferases in Mammalian Cells. Methods Mol Biol *1766*, 157-174.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Martinez-Rucobo, F.W., and Cramer, P. (2013). Structural basis of transcription elongation. Biochim Biophys Acta *1829*, 9-19.

Mason, P.B., and Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol *23*, 8323-8333.

Mattiroli, F., D'Arcy, S., and Luger, K. (2015). The right place at the right time: chaperoning core histone variants. EMBO Rep *16*, 1454-1466.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K.*, et al.* (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res *34*, D108-110.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y.*, et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253-257.

Maurano, M.T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R., and Stamatoyannopoulos, J.A. (2015). Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. Nat Genet *47*, 1393-1401.

Maze, I., Wenderski, W., Noh, K.M., Bagot, R.C., Tzavaras, N., Purushothaman, I., Elsasser, S.J., Guo, Y., Ionete, C., Hurd, Y.L.*, et al.* (2015). Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. Neuron *87*, 77-94.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179-1186.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

McDaniel, S.L., and Strahl, B.D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. Cell Mol Life Sci.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A *101*, 1525-1530.

Meers, M.P., Henriques, T., Lavender, C.A., McKay, D.J., Strahl, B.D., Duronio, R.J., Adelman, K., and Matera, A.G. (2017). Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity. Elife *6*.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223-226.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Millar, A.H., Heazlewood, J.L., Giglione, C., Holdsworth, M.J., Bachmair, A., and Schulze, W.X. (2019). The Scope, Functions, and Dynamics of Posttranslational Protein Modifications. Annu Rev Plant Biol *70*, 119-151.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393-396.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755-766.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S.*, et al.* (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife *4*, e06205.

Muraro, M.J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.*, et al.* (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst *3*, 385-394 e383.

Musselman, C.A., Lalonde, M.E., Cote, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol *19*, 1218-1227.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Nicolas, D., Zoller, B., Suter, D.M., and Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. Proc Natl Acad Sci U S A *115*, 7153-7158.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526-540.

Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. Cell *149*, 1438-1445.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105-116.

Orsi, G.A., Couble, P., and Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. Int J Dev Biol *53*, 231-243.

Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. Mol Cell *3*, 379-387.

Papai, G., Frechard, A., Kolesnikova, O., Crucifix, C., Schultz, P., and Ben-Shem, A. (2020). Structure of SAGA and mechanism of TBP deposition on gene promoters. Nature *577*, 711-716.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., Hacker, K.E.*, et al.* (2016a). Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell *166*, Pages 950–962.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.C., Davis, I.J., Fahey, C.C., Hacker, K.E.*, et al.* (2016b). Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell *166*, 950-962.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J *23*, 4061-4071.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N.*, et al.* (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res *38*, 4958-4969.

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am J Hum Genet *31*, 106-115.

Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol *10*, 886-895.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.

Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., La Thangue, N.B., Porter, A.C.*, et al.* (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell Rep *7*, 2006-2018.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Pillutla, R.C., Yue, Z., Maldonado, E., and Shatkin, A.J. (1998). Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes. J Biol Chem *273*, 21443-21446.

Ping, X.L., Sun, B.F., Wang, L., Xiao, W., Yang, X., Wang, W.J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.S.*, et al.* (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res *24*, 177-189.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet *32*, 508-522.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet *8*, e1002717.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Pu, M., Ni, Z., Wang, M., Wang, X., Wood, J.G., Helfand, S.L., Yu, H., and Lee, S.S. (2015). Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. Genes Dev *29*, 718-731.

Rae, M.M., and Franke, W.W. (1972). The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. Chromosoma *39*, 443-456.

Rainier, S., and Feinberg, A.P. (1994). Genomic imprinting, DNA methylation, and cancer. J Natl Cancer Inst *86*, 753-759.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D.*, et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H.*, et al.* (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol *14*, R25.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr Opin Genet Dev *43*, 73-81.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J *31*, 593-605.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol *194*, 357-367.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C., and Ferrara, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal Biochem *203*, 173-179.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. Genes Dev *20*, 1050-1056.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187-1200.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J.*, et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237-240.

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., Yamagiwa, A., Niwa, H., Wakayama, T., and Okano, M. (2010). DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. Curr Biol *20*, 1452-1457.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Scharer, O.D. (2013). Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol *5*, a012609.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34-57.

Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. Nat Rev Mol Cell Biol *19*, 526-541.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev *16*, 919-932.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Scialdone, A., Natarajan, K.N., Saraiva, L.R., Proserpio, V., Teichmann, S.A., Stegle, O., Marioni, J.C., and Buettner, F. (2015). Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods *85*, 54-61.

Sen, P., Dang, W., Donahue, G., Dai, J., Dorsey, J., Cao, X., Liu, W., Cao, K., Perry, R., Lee, J.Y.*, et al.* (2015). H3K36 methylation promotes longevity by enhancing transcriptional fidelity. Genes Dev *29*, 1362-1376.

Shandilya, J., and Roberts, S.G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. Biochim Biophys Acta *1819*, 391-400.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645-653.

Sheridan, R.M., Fong, N., D'Alessandro, A., and Bentley, D.L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. Mol Cell *73*, 107-118 e104.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T.*, et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Sigurdsson, S., Dirac-Svejstrup, A.B., and Svejstrup, J.Q. (2010). Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. Mol Cell *38*, 202-210.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697-708.

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.F., Jonasch, E., Furey, T.S.*, et al.* (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res *24*, 241-250.

Skucha, A., Ebner, J., Schmollerl, J., Roth, M., Eder, T., Cesar-Razquin, A., Stukalov, A., Vittori, S., Muhar, M., Lu, B.*, et al.* (2018). MLL-fusion-driven leukemia requires SETD2 to safeguard genomic integrity. Nat Commun *9*, 1983.

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. Elife *5*.

Sleeth, K.M., Sorensen, C.S., Issaeva, N., Dziegielewski, J., Bartek, J., and Helleday, T. (2007). RPA mediates recombination repair during replication stress and is displaced from DNA by checkpoint signalling in human cells. J Mol Biol *373*, 38-47.

Slobodin, B., Han, R., Calderone, V., Vrielink, J., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326-337 e312.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204-220.

Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A *82*, 6470-6474.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G.A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO J *21*, 1764-1774.

Stewart, M. (2019). Polyadenylation and nuclear export of mRNAs. J Biol Chem *294*, 2977-2987.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R.*, et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L.*, et al.* (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. Mol Cell *70*, 371-379 e375.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol *14*, 103-105.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213-1226.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tonjes, M., Sill, M., Bender, S.*, et al.* (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell *22*, 425-437.

Sun, X.J., Wei, J., Wu, X.Y., Hu, M., Wang, L., Wang, H.H., Zhang, Q.H., Chen, S.J., Huang, Q.H., and Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem *280*, 35261-35271.

Surova, O., and Zhivotovsky, B. (2013). Various modes of cell death induced by DNA damage. Oncogene *32*, 3789-3797.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P.*, et al.* (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res *47*, D607-D613.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H.*, et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev *16*, 1779-1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev *19*, 815-826.

Teissandier, A., and Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. EMBO J *36*, 1471-1473.

Teloni, F., Michelena, J., Lezaja, A., Kilic, S., Ambrosi, C., Menon, S., Dobrovolna, J., Imhof, R., Janscak, P., Baubec, T.*, et al.* (2019). Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability. Mol Cell *73*, 670-683 e612.

Terzo, E.A., Lim, A.R., Chytil, A., Chiang, Y.C., Farmer, L., Gessner, K.H., Walker, C.L., Jansen, V.M., and Rathmell, W.K. (2019). SETD2 loss sensitizes cells to PI3Kbeta and AKT inhibition. Oncotarget *10*, 647-659.

Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem Sci *39*, 577-586.

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A *98*, 8614-8619.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res *13*, 2129-2141.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Tian, B., Yang, J., and Brasier, A.R. (2012). Two-step cross-linking for analysis of protein-chromatin interactions. Methods Mol Biol *809*, 105-120.

Tiedemann, R.L., Hlady, R.A., Hanavan, P.D., Lake, D.F., Tibes, R., Lee, J.H., Choi, J.H., Ho, T.H., and Robertson, K.D. (2015). Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. Oncotarget.

Tiedemann, R.L., Hlady, R.A., Hanavan, P.D., Lake, D.F., Tibes, R., Lee, J.H., Choi, J.H., Ho, T.H., and Robertson, K.D. (2016). Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. Oncotarget *7*, 1927-1946.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol *16*, 996-1001.

Toledo, L.I., Altmeyer, M., Rask, M.B., Lukas, C., Larsen, D.H., Povlsen, L.K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. Cell *155*, 1088-1103.

Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L., and Hurley, J.H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. Cell *111*, 91-103.

Tsai, K.L., Yu, X., Gopalan, S., Chao, T.C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W.*, et al.* (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature *544*, 196-201.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Turinetto, V., and Giachino, C. (2015). Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. Nucleic Acids Res *43*, 2489-2498.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.

van Nimwegen, E. (2003). Scaling laws in the functional content of genomes. Trends Genet *19*, 479-484.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature *489*, 452-455.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G.*, et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Villasenor, R., Pfaendler, R., Ambrosi, C., Butz, S., Giuliani, S., Bryan, E., Sheahan, T.W., Gable, A.L., Schmolka, N., Manzo, M.*, et al.* (2020). ChromID identifies the protein interactome at chromatin marks. Nat Biotechnol.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. Mol Cell *75*, 310-323 e318.

Vo, N., and Goodman, R.H. (2001). CREB-binding protein and p300 in transcriptional regulation. J Biol Chem *276*, 13505-13508.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

Vos, S.M., Pollmann, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F., and Cramer, P. (2016). Architecture and RNA binding of the human negative elongation factor. Elife *5*.

Waddington, C.H. (1957). The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology (London: Allen & Unwin).

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol *13*, 115-126.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Wang, W.Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., Mackenzie, I.R., Huang, E.J., and Tsai, L.H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat Neurosci *16*, 1383-1391.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G.*, et al.* (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117-120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell *161*, 1388-1399.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet *37*, 853-862.

Wei, X., Samarabandu, J., Devdhar, R.S., Siegel, A.J., Acharya, R., and Berezney, R. (1998). Segregation of transcription and replication sites into higher order domains. Science *281*, 1502-1506.

Weinberger, L., Voichek, Y., Tirosh, I., Hornung, G., Amit, I., and Barkai, N. (2012). Expression noise and acetylation profiles distinguish HDAC functions. Mol Cell *47*, 193-202.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-140.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J.*, et al.* (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature *508*, 263-268.

Wen, Y., Yue, Z., and Shatkin, A.J. (1998). Mammalian capping enzyme binds RNA and uses protein tyrosine phosphatase mechanism. Proc Natl Acad Sci U S A *95*, 12226-12231.

Williams, A.S., and Marzluff, W.F. (1995). The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. Nucleic Acids Res *23*, 654-662.

Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. Genes Dev *19*, 1761-1766.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell *54*, 601-612.

Wood, A., and Shilatifard, A. (2004). Posttranslational modifications of histones by methylation. Adv Protein Chem *67*, 201-222.

Wright, W.D., Shah, S.S., and Heyer, W.D. (2018). Homologous recombination and the repair of DNA double-strand breaks. J Biol Chem *293*, 10524-10535.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev *17*, 1402-1414.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y.*, et al.* (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Mol Cell *61*, 507-519.

Xie, L., Pelz, C., Wang, W., Bashar, A., Varlamova, O., Shadle, S., and Impey, S. (2011). KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J *30*, 1473-1484.

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., Xia, W.*, et al.* (2019). SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nat Genet *51*, 844-856.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Young, R.A. (1991). RNA polymerase II. Annu Rev Biochem *60*, 689-715.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D.*, et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat Cell Biol *16*, 2-9.

Zhang, H., Rigo, F., and Martinson, H.G. (2015a). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol Cell *59*, 437-448.

Zhang, T., Cooper, S., and Brockdorff, N. (2015b). The interplay of histone modifications - writers that read. EMBO Rep *16*, 1467-1481.

Zhang, Y., Xie, S., Zhou, Y., Xie, Y., Liu, P., Sun, M., Xiao, H., Jin, Y., Sun, X., Chen, Z.*, et al.* (2014). H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. Cell Rep *8*, 1989-2002.

Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol *18*, 31-42.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Z., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m(6)A through the Low-Complexity Protein hnRNPG. Mol Cell *76*, 70-81 e79.

Zhu, K., Lei, P.J., Ju, L.G., Wang, X., Huang, K., Yang, B., Shao, C., Zhu, Y., Wei, G., Fu, X.D.*, et al.* (2017). SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. Nucleic Acids Res *45*, 92-105.

10. Curriculum Vitae

Personal Information

NAME	Christina Ambrosi
ADDRESS	Schwamendingenstrasse 54, CH-8050 Zurich
TELEPHONE	+41 76 332 0407
EMAIL	christina.ambrosi@uzh.ch
NATIONALITY	German
DATE OF BIRTH	04.07.1990
HOME TOWN	Berlin, Germany

Education

SINCE 11/2015	**Molecular Life Science PhD Program, Zurich** University of Zurich, Switzerland PhD Thesis: *"Exploring the Functional Role of SETD2/H3K36me3 during Cellular Differentiation"*
10/2013 - 09/2015	**Master of Science, Molecular Medicine** Charité Universitätsmedizin Berlin, Germany Master Thesis: *"Modification of the SOX9 locus in murine embryonic stem cells by the CRISPR/Cas9"*
10/2010 - 09/2013	**Bachelor of Science, Biology** Free University Berlin, Germany Bachelor Thesis: *"Funktionelle Charakterisierung einer β-Catenin-Mutante im Rahmen des Androgenrezeptorsignalwegs"*

Awards

2018	Funding by Forschungskredit Candoc, University of Zurich
2017	Travel Grant, Molecular Life Sciences PhD program, LSZGS

Languages

German	mother tongue
English	level C2
French	level A1
Italian	level A1

Publications

ORCID-ID: 0000-0002-7167-1173

- *ChromID identifies the protein interactome at chromatin marks*; Villaseñor R, Pfaendler R, **Ambrosi C**, Butz S, Giuliani S, Bryan E, Sheahan TW, Gable AL, Schmolka N, Manzo M, Wirz J, Feller C, von Mering C, Aebersold R, Voigt P and Baubec T; Nature Biotechnology, 2020
- *Efficient pre-mRNA cleavage prevents replication stress-associated genome instability*; Teloni F, Michelena J, Lezaja A, Kilic S, **Ambrosi C**, Menon S, Dobrovolna J, Imhof R, Janscak P, Baubec T, and Altmeyer M; Molecular Cell, 2019
- *Genome-wide profiling of DNA methyltransferases in mammalian cells* (Method); Manzo M, **Ambrosi C** and Baubec T; Springer Methods in Molecular Biology, 2018
- *Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands*; Manzo M, Wirz J, **Ambrosi C**, Villaseñor R, Roschitzki B & Baubec T; The EMBO Journal, 2017
- *Dynamics and Context-Dependent Roles of DNA Methylation* (Review); **Ambrosi C**, Manzo M and Baubec T; Journal of Molecular Biology, 2017